STIC-Biotech/ChemLib 88592

From: Schnizer, Richard
Sent: Monday, March 10, 2003 12:11 PM
To: STIC-Biotech/ChemLib
Subject: 09/836,439

Please search the commercial databases for SEQ ID NOS: 1-6 from 09/836,439.

Thank you-

Richard Schnizer, Ph.D.
Patent Examiner
Art Unit 1635
CM1 12E17
703-306-5441
Mail Box CM1 11E12

CM1 12E

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

	TYPE OF SEARCH:	VENDOR/COST (where applic.)
Searcher:_____	NA Sequences:_____	STN:_____
Phone:_____	AA Sequences:_____	DIALOG:_____
Location:_____	Structures:_____	Questel/Orbit:_____
Date Picked Up: 3/11	Bibliographic:_____	DRLink:_____
Date Completed: 3/19	Litigation:_____	Lexis/Nexis:_____
Searcher Prep/Review:_____	Full text:_____	Sequence Sys.:_____
Clerical:_____	Patent Family:_____	WWW/Internet:_____
Online time:_____	Other:_____	Other (specify):_____

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:43:42 ; Search time 625.892 Seconds
(without alignments)
3161.870 Million cell updates/sec

Title: US-09-836-439-1
Perfect score: 68
Sequence: 1 ccttccaacctacgtagcag.....ggaagggcgcgttttcgcgc 68

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	28	41.2	68	6	AR024440	AR024440 Sequence
2	28	41.2	68	6	I93657	I93657 Sequence 23
c 3	27.2	40.0	10591	2	AC020132	AC020132 Drosophil
4	27.2	40.0	73076	3	AC003055	AC003055 Drosophil
c 5	27.2	40.0	168540	3	AC092189	AC092189 Drosophil
c 6	27.2	40.0	303356	3	AE003583	AE003583 Drosophil
7	27	39.7	3554	3	DMFTZ1	X00854 Drosophila
8	27	39.7	56043	2	AC012761	AC012761 Drosophil
9	27	39.7	66991	3	AC001653	AC001653 Drosophil
c 10	27	39.7	74026	2	AC101196	AC101196 Mus muscu
11	27	39.7	80866	2	AC012649	AC012649 Drosophil
c 12	27	39.7	110000	3	AE001572_1	Continuation (2 of
13	27	39.7	190642	3	AC095015	AC095015 Drosophil
14	27	39.7	208375	2	AC107703	AC107703 Mus muscu
15	27	39.7	309357	3	AE003673	AE003673 Drosophil
c 16	26.6	39.1	91104	2	AC112364	AC112364 Rattus no
c 17	26.6	39.1	160589	10	AL663066	AL663066 Mouse DNA
18	26.4	38.8	68	6	AR024439	AR024439 Sequence
19	26.4	38.8	68	6	I93656	I93656 Sequence 22
20	26.4	38.8	20281	1	AE008861	AE008861 Salmonell
21	26.4	38.8	192675	10	AL589767	AL589767 Mouse DNA
22	26.4	38.8	199842	2	AC094068	AC094068 Rattus no
23	26.4	38.8	207922	10	AL450399	AL450399 Mouse DNA
24	26.2	38.5	139877	9	AC004066	AC004066 Homo sapi
c 25	26.2	38.5	194521	2	AC093696	AC093696 Homo sapi
c 26	26	38.2	161719	2	AC109157	AC109157 Mus muscu
27	26	38.2	183311	2	AC109788	AC109788 Bos tauru
28	26	38.2	187252	10	AL607109	AL607109 Mouse DNA
c 29	26	38.2	222941	2	AC102236	AC102236 Mus muscu
c 30	25.8	37.9	2693	10	MMPB	X60133 Murine MPB
c 31	25.8	37.9	2703	10	MMPBALT	X87952 Murine MPB
c 32	25.8	37.9	2743	10	MMPDE	X55968 Mouse mRNA
33	25.8	37.9	129790	3	AGA441131	AJ441131 Anopheles
c 34	25.8	37.9	157418	2	AC006912	AC006912 Caenorhab
c 35	25.8	37.9	163475	9	AC092619	AC092619 Homo sapi
36	25.6	37.6	154552	9	AL390964	AL390964 Human DNA
c 37	25.6	37.6	190721	8	AP003263	AP003263 Oryza sat
38	25.4	37.4	435	6	AX341469	AX341469 Sequence
39	25.4	37.4	1224	9	BC007520	BC007520 Homo sapi
40	25.4	37.4	1324	9	AK026222	AK026222 Homo sapi
41	25.4	37.4	2727	6	AX282833	AX282833 Sequence
42	25.4	37.4	2730	6	AX282845	AX282845 Sequence
43	25.4	37.4	2874	6	AX282837	AX282837 Sequence
44	25.4	37.4	2877	6	AX282849	AX282849 Sequence
45	25.4	37.4	2979	6	AX282841	AX282841 Sequence

ALIGNMENTS

RESULT 1
AR024440
LOCUS AR024440 68 bp DNA linear PAT 05-DEC-1998
DEFINITION Sequence 23 from patent US 5795972.
ACCESSION AR024440
VERSION AR024440.1 GI:3977734
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 68)
AUTHORS Kmiec,E.B.
TITLE Chimeric mutational vectors having non-natural nucleotides
JOURNAL Patent: US 5795972-A 23 18-AUG-1998;
FEATURES Location/Qualifiers

```
     source          1. .68
                     /organism="unknown"
BASE COUNT       11 a     19 c     19 g     19 t
ORIGIN

  Query Match             41.2%;  Score 28;  DB 6;  Length 68;
  Best Local Similarity   61.7%;  Pred. No. 4.3;
  Matches   37;  Conservative    3;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       9 CCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCGCGC 68
           |||  | || |||  | |||| |  :|: |: |  |||    | | ||||||||||||||
Db       9 CCTGAGGAGAAGACTGCTTTTGCAGTCTTCTCCTCAGGAGTCAGGTGCGCGTTTTCGCGC 68


RESULT 2
I93657
LOCUS       I93657                    68 bp    DNA     linear   PAT 01-DEC-1998
DEFINITION  Sequence 23 from patent US 5731181.
ACCESSION   I93657
VERSION     I93657.1  GI:3938127
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68)
  AUTHORS   Kmiec,E.B.
  TITLE     Chimeric mutational vectors having non-natural nucleotides
  JOURNAL   Patent: US 5731181-A 23 24-MAR-1998;
FEATURES             Location/Qualifiers
     source          1. .68
                     /organism="unknown"
BASE COUNT       11 a     19 c     19 g     19 t
ORIGIN

  Query Match             41.2%;  Score 28;  DB 6;  Length 68;
  Best Local Similarity   61.7%;  Pred. No. 4.3;
  Matches   37;  Conservative    3;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       9 CCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCGCGC 68
           |||  | || |||  | |||| |  :|: |: |  |||    | | ||||||||||||||
Db       9 CCTGAGGAGAAGACTGCTTTTGCAGTCTTCTCCTCAGGAGTCAGGTGCGCGTTTTCGCGC 68


RESULT 3
AC020132/c
LOCUS       AC020132               10591 bp    DNA     linear   HTG 03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC020132
VERSION     AC020132.1  GI:6664765
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 10591)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212153 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .10591
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     2920 a   2216 c   2177 g   3278 t
ORIGIN

  Query Match             40.0%;  Score 27.2;  DB 2;  Length 10591;
  Best Local Similarity   60.7%;  Pred. No. 12;
  Matches   34;  Conservative    4;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      11 TACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCGC 66
           ||||| || ||||||||||  :: |:|| | |||  :  ||| |     ||||| |
Db    3352 TACGTCGCTGAAAGTTTTTCGTTCCTGCCATGTAATTACGAACGCATTCTTTTCAC 3297


RESULT 4
AC003055
LOCUS       AC003055               73076 bp    DNA     linear   INV 04-NOV-1997
DEFINITION  Drosophila melanogaster (P1 DS06332 (D91)) DNA sequence, complete
            sequence.
ACCESSION   AC003055 AC001994 AC002947 AC002944 AC002943 AC002952 AC001992
            AC002945 AC002946 AC001988 AC001995 AC002951 AC001991 AC002942
            AC001993 AC002948 AC002953 AC001990 AC001989 AC002949 AC002950
VERSION     AC003055.1  GI:2584828
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in pOT2 from P1 clone DS06332
            (D91)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 73076)
  AUTHORS   Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Doyle,C.M.,
            Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Kim,S.H., Ko,C.L., Li,M., Lomotan,M.A., Mazda,P.,
            Mok,M.S., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D.,
            Santos,R.F., Snir,E., Stevko,V., Subramanian,S., Towne,B.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and
            Kimmel,B.
  TITLE     Sequencing of Drosophila chromosome 2L, region 22F
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 73076)
  AUTHORS   Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Doyle,C.M.,
            Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Kim,S.H., Ko,C.L., Li,M., Lomotan,M.A., Mazda,P.,
            Mok,M.S., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D.,
            Santos,R.F., Snir,E., Stevko,V., Subramanian,S., Towne,B.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and
            Kimmel,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1997) Berkeley Drosophila Genome Project, MS
            74-157, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 74-157
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to  other sequences, please visit our sequence
            archive Web site (http://www-hgc.lbl.gov/sequence-archive.html) or
            send email to drosophila@genome.lbl.gov.
            Library location: 66-92.
            This P1 was assembled from the following subclones: 2_c1
            (AC001994), 2_b12 (AC002947), 1_g6 (AC002944), 1_d4 (AC002943),
            2_b5, 2_e4 (AC002952), 1_b4, 2_a1, 2_a2, 1_h6 (AC001992), 1_h8
            (AC002945), 2_a8 (AC002946), 1_b2 (AC001988), 2_c5 (AC001995), 2_d5
            (AC002951), 1_h3 (AC001991), 1_d3 (AC002942), 2_a9 (AC001993),
            2_c10 (AC002948), 2_g9 (AC002953), 1_g7 (AC001990), 1_b5
            (AC001989), 2_c3 (AC002949), 2_h2, 2_c4 (AC002950), 2_b6.
FEATURES             Location/Qualifiers
     source          1. .73076
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="22F2-22F4"
                     /clone="P1 DS06332 (D91)"
BASE COUNT     21203 a  14890 c  15769 g  21214 t
ORIGIN
```

```
  Query Match             40.0%;  Score 27.2;  DB 3;  Length 73076;
  Best Local Similarity   60.7%;  Pred. No. 14;
  Matches   34;  Conservative    4;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      11 TACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCGC 66
           ||||| || ||||||||||  :: |:|| | |||  :  ||| |     ||||| |
Db   61972 TACGTCGCTGAAAGTTTTTCGTTCCTGCCATGTAATTACGAACGCATTCTTTTCAC 62027


RESULT 5
AC092189/c
LOCUS       AC092189              168540 bp    DNA     linear   INV 06-SEP-2001
DEFINITION  Drosophila melanogaster, chromosome 2L, region 22E-22F, BAC clone
            BACR34M11, complete sequence.
ACCESSION   AC092189
VERSION     AC092189.2  GI:15451489
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 168540)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2L, region 22E-22F
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168540)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2001) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     On Sep 6, 2001 this sequence version replaced gi:14550291.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .168540
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="22E-22F"
                     /clone="BACR34M11 (D1362)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    47514 a  37192 c  36097 g  47737 t
ORIGIN

  Query Match             40.0%;  Score 27.2;  DB 3;  Length 168540;
  Best Local Similarity   60.7%;  Pred. No. 15;
  Matches   34;  Conservative    4;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      11 TACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCGC 66
           ||||| || ||||||||||  :: |:|| | |||  :  ||| |     ||||| |
Db   93783 TACGTCGCTGAAAGTTTTTCGTTCCTGCCATGTAATTACGAACGCATTCTTTTCAC 93728


RESULT 6
AE003583/c
LOCUS       AE003583              303356 bp    DNA     linear   INV 04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386046 section 8
            of 16, complete sequence.
ACCESSION   AE003583 AE002638
VERSION     AE003583.2  GI:10727375
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 303356)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
```

```
   PUBMED   10731132
REFERENCE   2  (bases 1 to 303356)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7295946.
FEATURES             Location/Qualifiers
     source          1. .303356
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
     gene            complement(<971. .>5749)
                     /gene="CG10882"
                     /map="22D2-22D5"
                     /db_xref="FLYBASE:FBan0010882"
                     /db_xref="FLYBASE:FBgn0031408"
     mRNA            complement(join(971. .1111,1167. .1970,2033. .2900,
                     3158. .3273,3332. .3791,3851. .4822,5231. .5749))
                     /gene="CG10882"
                     /product="CT30485"
                     /db_xref="FLYBASE:FBan0010882"
                     /db_xref="FLYBASE:FBgn0031408"
     CDS             complement(join(987. .1111,1167. .1970,2033. .2900,
                     3158. .3273,3332. .3791,3851. .4822,5231. .5464))
                     /gene="CG10882"
                     /note="CG10882 gene product"
                     /codon_start=1
                     /protein_id="AAF51283.1"
                     /db_xref="GI:7295985"
                     /db_xref="FLYBASE:FBan0010882"
                     /db_xref="FLYBASE:FBgn0031408"
                     /translation="MNPNMYGPPPTQFQQQPQFGAPPPNSGGWPPQQQQLPQQQPPQQ
                     QLPPQQQQQQPQYGAPPPTSAASQPYLNGNYQQQLATSMGGLSVGGGVGGANPLKPPL
                     PQGAPAAAAPPPTGFNQFNSNAAPPPTNNNNAAFGAPPPTQAGSYVNGALPPSSTPQS
                     VASGINQMSLNSATLAGLPHMPPPKAATPGAAPGQPPIPAAGSTSQPPLPGQPPLPGQ
                     PPFSGQIPTSQPAPSPYGVPSSRPGQPQLPPGATPPTYTQPGLPPQQQQGIPPLQQPG
                     IPQQQPGFPPQQPGLPPLSQPGLPPQPGAPYGAPQQGGYSGGFPGQAPGGFPGAPPPL
                     PGQQAAAPPQFGAPQPGYPGQQPGYPPQPGQQPMPGYPPQPGQQLGGPGYPPQPGAGF
                     PGQPGRPGFNQPPMPGAGNMYQQAPQARRLDPDQMPNPIQVMIENQRLSGGPFVTNQP
                     GLLPPLVTTKFVVHDQGNSSPRFLRSSLYCIPNTGDLLKTTALPLTLNISPLAKVGEG
                     EMEPPIVNFGEMGPIRCNRCKAYMSPNMQFVDAGRRFQCLMCKVTSEVHQNYYQHLDH
                     TGQRVDKHERPELLLGTYEFLATKDYCRNNTPPEVPAFIFIIDVSYNTVKSGLVHLLC
                     SQIKNILKHLPVDQGQDKSKVRVGFITYNSTVHFYNIKSSLAQPQMMVVGDVQEMFMP
                     LLDGFLCHPEESAAVIDALMEEIPRMFADTKETETILYPAIQAGLEALKASNAAGKLL
                     VFNSTLPIAEAPGKLKNRDDRKLLGTDKEKTVLTPQTTAYNTLGQECVQQGCSVDLFV
                     FNNAYIDLATIGQVSRLTGGEVFKYTYFQADVDGKRLIQDIIKNVSRPIAFDAVMRVR
                     TSAGIRPTEFYGHFFMSNTTDVELASIDATKSISIEIKHDDKLAPEENVYLQVALLYT
                     SCSGQRRLRILNLALRVTTTIADVFKCCDLDAMMLFFAKQACFKLMEHSPKQVKDNLI
                     HRSAQILACYRKHCTSPTSAGQLILPECLKLLPLYASCLLKNDAISGGSDMTLDDRSY
                     VIQFILSMDLNQSVSYLYPRFIPIHNVVPEETDLPTPVRCTHEKTQEDGAYILENGVH
                     LFVWLGQALSPDFVQSVFGVQGLQQIALERFNIVPETPLAKRIHGILEQIMKERSRYM
                     RVRICNNLVAAERQTGERVPTLPRGGPRYRRFRQLRGFPVSHAQGD"
     gene            complement(<6055. .>9191)
                     /gene="CG17006"
                     /map="22D2-22D3"
                     /db_xref="FLYBASE:FBan0017006"
                     /db_xref="FLYBASE:FBgn0031404"
     mRNA            complement(join(6055. .6897,6960. .7054,7301. .7841,
                     7911. .8051,8115. .8293,8640. .9191))
                     /gene="CG17006"
                     /product="CT37745"
                     /db_xref="FLYBASE:FBan0017006"
                     /db_xref="FLYBASE:FBgn0031404"
     CDS             complement(join(8119. .8298,8640. .9113))
                     /gene="CG17006"
                     /note="CG17006 gene product"
                     /codon_start=1
                     /protein_id="AAF51282.1"
                     /db_xref="GI:7295984"
                     /db_xref="FLYBASE:FBan0017006"
                     /db_xref="FLYBASE:FBgn0031404"
                     /translation="MKYGFPSTNDITINETFDFVTSFDRRNSAILWMCERVDLSNRVV
                     YGDSTSVAPAGAFGQSEAARVFFLSNIRPFLNRGFNLTVWDRLLQYVHEMSQRHGTVY
                     AYTGSIYLPRELKSNSWFLEFQSEERTMVAVPTHFFKILVIDKKFAGDTIPYAEAYSP
                     FHGPKMAAPGTGFRPGSDCALCILRGWNVFSAHRCQKTPAGTNTNGSLCLLSQSQNL"
     gene            <11323. .>13338
                     /gene="CG4267"
                     /map="22D2-22D3"
                     /db_xref="FLYBASE:FBan0004267"
                     /db_xref="FLYBASE:FBgn0031405"
     mRNA            join(11323. .11514,11616. .11868,12483. .13338)
                     /gene="CG4267"
                     /product="CT11345"
                     /db_xref="FLYBASE:FBan0004267"
                     /db_xref="FLYBASE:FBgn0031405"
     CDS             join(11445. .11514,11616. .11868,12483. .13284)
                     /gene="CG4267"
                     /note="CG4267 gene product"
                     /codon_start=1
                     /protein_id="AAF51281.1"
                     /db_xref="GI:7295983"
                     /db_xref="FLYBASE:FBan0004267"
                     /db_xref="FLYBASE:FBgn0031405"
                     /translation="MISQLGRISIMSAVLLAFAAIVFASSSEKTELDPNATCNYTLVK
                     KKTLGVDPSFWKKFFKHLIPFTSSRGKMQFILFKRDFADCGRELFVGDVENLRNSGFD
                     ARHQTRIVIHGWMSQSKGSHIRKVKNAYLSLTDPGPNGEPAPYEDFNVIVCDWSKTST
                     NVNYYEVAKTVEDMGALLAELVRYLNQEANMHYDDVYVIGHSLGAQIAGSAGKQIMPY
                     RFNTIYALDPAGPQFREKSDEYRIDASDASYVESIQTSVSFGFEQPVGHATFYPNYGK
                     NQKKCYVYGCSHKRSHDYFIESLTSPAGFWGPRCERHDDGTWLLLMSDGEFRMGGEPS
                     IPKNGTFYVKTYSKPPYAMGHRWQTEPPPREDDIENSTEE"
     gene            complement(<19456. .>20193)
                     /gene="Ser12"
                     /note="CG17240"
                     /evidence=not_experimental
                     /map="22D3-22D4"
                     /db_xref="FLYBASE:FBan0017240"
                     /db_xref="FLYBASE:FBgn0011832"
     mRNA            complement(<19456. .>20193)
                     /gene="Ser12"
                     /product="CT30487"
                     /note="Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /evidence=not_experimental
                     /db_xref="FLYBASE:FBan0017240"
                     /db_xref="FLYBASE:FBgn0011832"
     CDS             complement(19456. .20193)
                     /gene="Ser12"
                     /note="Ser12 gene product; Nucleotide sequence of the
                     Celera sequence differs from the published sequence for
                     this transcript"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAF51280.1"
                     /db_xref="GI:7295982"
                     /db_xref="FLYBASE:FBan0017240"
                     /db_xref="FLYBASE:FBgn0011832"
                     /translation="MLLHWLVLVASVTLISAGSSPERIVGGHPVLISEVPWQAALMYS
                     EKYICGAVIYSDKIIITAAHCVERPFDTLYSVRVGSVWKNLGGQHARVAVIRKHEDYV
                     SSTILFNDIAVIRLVDTLIFNAEVRPIQLADSAPAAGTEASVSGWGEIGILWLQPTSL
                     LKTSVKILDPNVCKRSYQYITKTMICAAALLKDSCHGDSGGPLVSGGQLVGIVSYGIG
                     CANPFFPGVYANVAELKPWILNAIEQL"
     gene            complement(<20507. .>21253)
                     /gene="CG17239"
                     /evidence=not_experimental
                     /db_xref="FLYBASE:FBan0017239"
                     /db_xref="FLYBASE:FBgn0042186"
     mRNA            complement(<20507. .>21253)
                     /gene="CG17239"
                     /product="CT30489"
                     /evidence=not_experimental
                     /db_xref="FLYBASE:FBan0017239"
                     /db_xref="FLYBASE:FBgn0042186"
     CDS             complement(20507. .21253)
                     /gene="CG17239"
                     /note="CG17239 gene product"
```

```
                     /codon_start=1

  Query Match             40.0%;  Score 27.2;  DB 3;  Length 303356;
  Best Local Similarity   60.7%;  Pred. No. 15;
  Matches   34;  Conservative    4;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      11 TACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCGC 66
           ||||| || |||||||||| :: |:|| | ||| :  ||| |     |||||| |
Db 277207 TACGTCGCTGAAAGTTTTTCGTTCCTGCCATGTAATTACGAACGCATTCTTTTCAC 277152


RESULT 7
DMFTZ1
LOCUS       DMFTZ1                  3554 bp    DNA     linear   INV 07-NOV-1997
DEFINITION  Drosophila melanogaster ftz gene.
ACCESSION   X00854 K01951
VERSION     X00854.1  GI:7984
KEYWORDS    DNA binding protein; inverted repeat; segmentation gene;
            unidentified reading frame.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 3554)
  AUTHORS   Laughon,A. and Scott,M.P.
  TITLE     Sequence of a Drosophila segmentation gene: protein structure
            homology with DNA-binding proteins
  JOURNAL   Nature 310 (5972), 25-31 (1984)
  MEDLINE   84245843
   PUBMED   6330566
REFERENCE   2  (bases 1 to 3554)
  AUTHORS   Scott,M.P. and Weiner,A.J.
  TITLE     Structural relationships among genes that control development:
            sequence homology between the Antennapedia, Ultrabithorax, and
            fushi tarazu loci of Drosophila
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 81 (13), 4115-4119 (1984)
  MEDLINE   84248068
   PUBMED   6330741
COMMENT     Sequence homology to Antp and Ubx genes; see DMANTP1 and DMANTP2
            sequence homology to prokaryotic and pot. yeast DNA-binding domains
            refer to Pabo C.O., Sauer R.T. (1984) A. Rev. Biochem. 53, 293-321.
            Data kindly reviewed (19-FEB-1986) by A. Laughon.
FEATURES             Location/Qualifiers
     source          1. .3554
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     TATA_signal     869. .875
     TATA_signal     880. .896
     prim_transcript 901. .2867
     exon            901. .1777
                     /number=1
     repeat_unit     901. .972
                     /note="inverted repeat"
     CDS             924. .992
                     /note="unidentified reading frame for pot. minor protein"
                     /codon_start=1
                     /protein_id="CAA25407.1"
                     /db_xref="GI:7985"
                     /translation="MQDLPQDQLIRKLTSVACTSQS"
     repeat_unit     979. .1048
                     /note="inverted repeat"
     CDS             join(1021. .1777,1928. .2412)
                     /codon_start=1
                     /product="ftz protein"
                     /protein_id="CAA25408.1"
                     /db_xref="GI:7986"
                     /db_xref="FLYBASE:FBgn0001077"
                     /db_xref="SWISS-PROT:P02835"
                     /translation="MATTNSQSHYSYADNMNMYNMYHPHSLPPTYYDNSGSNAYYQNT
                     SNYHSYQGYYPQESYSESCYYYNNQEQVTTQTVPPVQPTTPPPKATKRKAEDDAASII
                     AAVEERPSTLRALLTNPVKKLKYTPDYFYTTVEQVKKAPAVTTKVTASPAPSYDQEYV
                     TVPTPSASEDVDYLDVYSPQSQTQKLKNGDFATPPPTTPTSLPPLEGISTPPQSPGEK
                     SSSAVSQEINHRIVTAPNGAGDFNWSHIEETLASDCKDSKRTRQTYTRYQTLELEKEF
                     HFNRYITRRRRIDIANALSLSERQIKIWFQNRRMKSKKDRTLDSSPEHCGAGYTAMLP
                     PLEATSTATTGAPSVPVPMYHHHQTTAAYPAYSHSHSHGYGLLNDYPQQQTHQQYDAY
                     PQQYQQQCSYQQHPQDLYHLS"
     misc_feature    1066
                     /note="pot. alternate translation start site for ftz
                     protein"
     intron          1778. .1927
                     /number=1
     exon            1928. .2867
                     /number=2
     misc_feature    1933. .2121
                     /note="region with homology to Antp and Ubx genes"
     misc_feature    2023. .2121
                     /note="pot. DNA-binding domain of the ftz protein"
     variation       2041. .2043
                     /note="GCC (Ala) is GTC (Val) in temperature sensitive
                     allele ftz (f47ts)"
     variation       2112. .2142
                     /note="mutation ftz (Rpl) causing 10 novel amino acids and
                     a premature stop codon (2140-2142)"
     polyA_signal    2844. .2849
                     /note="pot. polyadenylation signal"
     polyA_signal    2848. .2854
                     /note="pot. polyadenylation signal"
     polyA_signal    2862. .2867
                     /note="pot. polyadenylation signal"
BASE COUNT      979 a    952 c    831 g    792 t
ORIGIN

  Query Match             39.7%;  Score 27;  DB 3;  Length 3554;
  Best Local Similarity   58.8%;  Pred. No. 14;
  Matches   30;  Conservative    6;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      15 TAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCG 65
           | ||| |||||||||||: : : |:|   |  ::|||||| |||  | || |
Db     171 TTGCATAAAGTTTTTACTGTTTACTAGTCATTTTGGAAGTGCGTTTGTTGG 221


RESULT 8
AC012761
LOCUS       AC012761                56043 bp    DNA     linear   HTG 03-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC012761
VERSION     AC012761.1  GI:6223082
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 56043)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210082 by the submitter.
            For further information on this sequence you may e-mail to
            fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .56043
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    16447 a  11385 c  11452 g  16759 t
ORIGIN

  Query Match             39.7%;  Score 27;  DB 2;  Length 56043;
  Best Local Similarity   58.8%;  Pred. No. 17;
```

```
  Matches   30;  Conservative    6;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      15 TAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCG 65
           | ||| |||||||||||: : : |:|   |  ::||||| |||  | || |
Db    5971 TTGCATAAAGTTTTTACTGTTTACTAGTCATTTTGGAAGTGCGTTTGTTGG 6021


RESULT 9
AC001653
LOCUS       AC001653               66991 bp    DNA     linear   INV 16-APR-1999
DEFINITION  Drosophila melanogaster, chromosome 3R, region 84B1-84B2, P1 clone
            DS07876, complete sequence.
ACCESSION   AC001653 L49396 L39779 L32657 L32652 L32645 L32656 L32649 L32648
            L32637 L39749 L32654 L32639 L32647 L32642 L32635 L32653 L32651
            L32638 L32655 L32640 L32641 L32646 L32650 L32636 L32643 L32644
VERSION     AC001653.1  GI:2342708
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 66991)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of antennapedia complex, homeotic genes
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 66991)
  AUTHORS   Martin,C.H., Arcaina,T.T., Bondoc,M.M., Chiang,A., Critz,P.A.,
            Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E., Gunning,K.M.,
            Houston,K.A., Jaklevic,M.A., Kadner,K.E., Kim,K., Kim,S.F.,
            Ko,C.L., Lewis,K.D., Li,M., Lindquist,K.J., Lomotan,M.A.,
            Lustre,V.M., Machrus,M.U., Mayeda,C.A., Miguel,T.M., Miller,C.A.,
            Mok,M.S., Pacleb,J.M., Patel,S.G., Santos,R.F., Subramanian,S.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C. and Palazzolo,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-APR-1997) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     On or before Apr 16, 1999 this sequence version replaced gi:483988,
            gi:483966, gi:1945589, gi:1103946.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
            P1 library location: 83-4.
FEATURES             Location/Qualifiers
     source          1. .66991
                     /organism="Drosophila melanogaster"
                     /strain="y2; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
                     /map="84B1-84B2"
                     /clone="P1 DS07876 (D14)"
                     /clone_lib="P1 library, partial Sau3A in pNS582tet14Ad10"
                     /note="This sequence has not changed since its original
                     submission on 08/25/1997.  It was resubmitted in order to
                     include all secondary accession numbers for the subclones
                     belonging to this clone."
BASE COUNT    19565 a  13581 c  13571 g  20274 t
ORIGIN
```

```
  Query Match           39.7%;  Score 27;  DB 3;  Length 66991;
  Best Local Similarity   58.8%;  Pred. No. 17;
  Matches   30;  Conservative    6;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      15 TAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCG 65
           | ||| |||||||||||: : : |:|   |  ::||||| |||  | || |
Db  25751 TTGCATAAAGTTTTTACTGTTTACTAGTCATTTTGGAAGTGCGTTTGTTGG 25801


RESULT 10
AC101196/c
LOCUS       AC101196               74026 bp    DNA     linear   HTG 23-NOV-2001
DEFINITION  Mus musculus clone RP23-177G5, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC101196
VERSION     AC101196.1  GI:17059970
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 74026)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-177G5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 74026)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L16139
                Center clone name: 177_G_5
            --------------.
            * NOTE: This record contains 93 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
```

```
*        1      705: contig of 705 bp in length
*      706 805: gap of      100 bp
*      806     1517: contig of 712 bp in length
*     1518 1617: gap of      100 bp
*     1618     2316: contig of 699 bp in length
*     2317 2416: gap of      100 bp
*     2417     3114: contig of 698 bp in length
*     3115 3214: gap of      100 bp
*     3215     3917: contig of 703 bp in length
*     3918 4017: gap of      100 bp
*     4018     4728: contig of 711 bp in length
*     4729 4828: gap of      100 bp
*     4829     5529: contig of 701 bp in length
*     5530 5629: gap of      100 bp
*     5630     6332: contig of 703 bp in length
*     6333 6432: gap of      100 bp
*     6433     7067: contig of 635 bp in length
*     7068 7167: gap of      100 bp
*     7168     7877: contig of 710 bp in length
*     7878 7977: gap of      100 bp
*     7978     8682: contig of 705 bp in length
*     8683 8782: gap of      100 bp
*     8783     9490: contig of 708 bp in length
*     9491 9590: gap of      100 bp
*     9591    10300: contig of 710 bp in length
*    10301 10400: gap of      100 bp
*    10401    11092: contig of 692 bp in length
*    11093 11192: gap of      100 bp
*    11193    11885: contig of 693 bp in length
*    11886 11985: gap of      100 bp
*    11986    12684: contig of 699 bp in length
*    12685 12784: gap of      100 bp
*    12785    13490: contig of 706 bp in length
*    13491 13590: gap of      100 bp
*    13591    14293: contig of 703 bp in length
*    14294 14393: gap of      100 bp
*    14394    15099: contig of 706 bp in length
*    15100 15199: gap of      100 bp
*    15200    15909: contig of 710 bp in length
*    15910 16009: gap of      100 bp
*    16010    16708: contig of 699 bp in length
*    16709 16808: gap of      100 bp
*    16809    17515: contig of 707 bp in length
*    17516 17615: gap of      100 bp
*    17616    18313: contig of 698 bp in length
*    18314 18413: gap of      100 bp
*    18414    19099: contig of 686 bp in length
*    19100 19199: gap of      100 bp
*    19200    19896: contig of 697 bp in length
*    19897 19996: gap of      100 bp
*    19997    20695: contig of 699 bp in length
*    20696 20795: gap of      100 bp
*    20796    21456: contig of 661 bp in length
*    21457 21556: gap of      100 bp
*    21557    22217: contig of 661 bp in length
*    22218 22317: gap of      100 bp
*    22318    23025: contig of 708 bp in length
*    23026 23125: gap of      100 bp
*    23126    23824: contig of 699 bp in length
*    23825 23924: gap of      100 bp
*    23925    24610: contig of 686 bp in length
*    24611 24710: gap of      100 bp
*    24711    25399: contig of 689 bp in length
*    25400 25499: gap of      100 bp
*    25500    26187: contig of 688 bp in length
*    26188 26287: gap of      100 bp
*    26288    26993: contig of 706 bp in length
*    26994 27093: gap of      100 bp
*    27094    27807: contig of 714 bp in length
*    27808 27907: gap of      100 bp
*    27908    28609: contig of 702 bp in length
*    28610 28709: gap of      100 bp
*    28710    29413: contig of 704 bp in length
*    29414 29513: gap of      100 bp
*    29514    30190: contig of 677 bp in length
*    30191 30290: gap of      100 bp
*    30291    30996: contig of 706 bp in length
*    30997 31096: gap of      100 bp
*    31097    31797: contig of 701 bp in length
*    31798 31897: gap of      100 bp
*    31898    32567: contig of 670 bp in length
*    32568 32667: gap of      100 bp
*    32668    33368: contig of 701 bp in length
*    33369 33468: gap of      100 bp
*    33469    34129: contig of 661 bp in length
*    34130 34229: gap of      100 bp
*    34230    34935: contig of 706 bp in length
*    34936 35035: gap of      100 bp
*    35036    35743: contig of 708 bp in length
*    35744 35843: gap of      100 bp
*    35844    36513: contig of 670 bp in length
*    36514 36613: gap of      100 bp
*    36614    37323: contig of 710 bp in length
*    37324 37423: gap of      100 bp
*    37424    38119: contig of 696 bp in length
*    38120 38219: gap of      100 bp
*    38220    38895: contig of 676 bp in length
*    38896 38995: gap of      100 bp
*    38996    39687: contig of 692 bp in length
*    39688 39787: gap of      100 bp
*    39788    40501: contig of 714 bp in length
*    40502 40601: gap of      100 bp
*    40602    41296: contig of 695 bp in length
*    41297 41396: gap of      100 bp
*    41397    42103: contig of 707 bp in length
*    42104 42203: gap of      100 bp
*    42204    42906: contig of 703 bp in length
*    42907 43006: gap of      100 bp
*    43007    43693: contig of 687 bp in length
*    43694 43793: gap of      100 bp
*    43794    44494: contig of 701 bp in length
*    44495 44594: gap of      100 bp
*    44595    45291: contig of 697 bp in length
*    45292 45391: gap of      100 bp
*    45392    46100: contig of 709 bp in length
*    46101 46200: gap of      100 bp
*    46201    46896: contig of 696 bp in length
*    46897 46996: gap of      100 bp
*    46997    47682: contig of 686 bp in length
*    47683 47782: gap of      100 bp
*    47783    48480: contig of 698 bp in length
*    48481 48580: gap of      100 bp
*    48581    49288: contig of 708 bp in length
*    49289 49388: gap of      100 bp
*    49389    50070: contig of 682 bp in length
*    50071 50170: gap of      100 bp
*    50171    50874: contig of 704 bp in length
*    50875 50974: gap of      100 bp
*    50975    51683: contig of 709 bp in length
*    51684 51783: gap of      100 bp
*    51784    52495: contig of 712 bp in length
*    52496 52595: gap of      100 bp
*    52596    53300: contig of 705 bp in length
*    53301 53400: gap of      100 bp
*    53401    54108: contig of 708 bp in length
*    54109 54208: gap of      100 bp
*    54209    54904: contig of 696 bp in length
*    54905 55004: gap of      100 bp

Query Match              39.7%;  Score 27;  DB 2;  Length 74026;
Best Local Similarity    67.4%;  Pred. No. 17;
Matches   29;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      21 AAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTT 63
           |||||| ||||:::| || | | ||| :| || ||| ||| ||
Db   29154 AAAGTTATTACTTTCAGCCAGGCAGGGTGAAAAGGCACGTATT 29112
```

RESULT 11
AC012649
LOCUS AC012649 80866 bp DNA linear HTG 04-NOV-1999
DEFINITION Drosophila melanogaster chromosome X clone BACR07N08 (D1121) RPCI-98 07.N.8 map 17D-17E strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 101 unordered pieces.
ACCESSION AC012649
VERSION AC012649.2 GI:6226990
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 80866)
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D., Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
TITLE Sequencing of Drosophila melanogaster
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 80866)
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Nov 4, 1999 this sequence version replaced gi:6175114.
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 101 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	577: contig of 577 bp in length
*	578	657: gap of unknown length
*	658	1404: contig of 747 bp in length
*	1405	1484: gap of unknown length
*	1485	2313: contig of 829 bp in length
*	2314	2393: gap of unknown length
*	2394	3131: contig of 738 bp in length
*	3132	3211: gap of unknown length
*	3212	3719: contig of 508 bp in length
*	3720	3799: gap of unknown length
*	3800	4897: contig of 1098 bp in length
*	4898	4977: gap of unknown length
*	4978	5799: contig of 822 bp in length
*	5800	5879: gap of unknown length
*	5880	6469: contig of 590 bp in length
*	6470	6549: gap of unknown length
*	6550	7442: contig of 893 bp in length
*	7443	7522: gap of unknown length
*	7523	8486: contig of 964 bp in length
*	8487	8566: gap of unknown length
*	8567	9246: contig of 680 bp in length
*	9247	9326: gap of unknown length
*	9327	9971: contig of 645 bp in length
*	9972	10051: gap of unknown length
*	10052	10632: contig of 581 bp in length
*	10633	10712: gap of unknown length
*	10713	11328: contig of 616 bp in length
*	11329	11408: gap of unknown length
*	11409	12070: contig of 662 bp in length
*	12071	12150: gap of unknown length
*	12151	12954: contig of 804 bp in length
*	12955	13034: gap of unknown length
*	13035	13887: contig of 853 bp in length
*	13888	13967: gap of unknown length
*	13968	14675: contig of 708 bp in length
*	14676	14755: gap of unknown length
*	14756	15777: contig of 1022 bp in length
*	15778	15857: gap of unknown length
*	15858	16561: contig of 704 bp in length
*	16562	16641: gap of unknown length
*	16642	17656: contig of 1015 bp in length
*	17657	17736: gap of unknown length
*	17737	18347: contig of 611 bp in length
*	18348	18427: gap of unknown length
*	18428	19269: contig of 842 bp in length
*	19270	19349: gap of unknown length
*	19350	20123: contig of 774 bp in length
*	20124	20203: gap of unknown length
*	20204	21037: contig of 834 bp in length
*	21038	21117: gap of unknown length
*	21118	21880: contig of 763 bp in length
*	21881	21960: gap of unknown length
*	21961	23063: contig of 1103 bp in length
*	23064	23143: gap of unknown length
*	23144	23742: contig of 599 bp in length
*	23743	23822: gap of unknown length
*	23823	24697: contig of 875 bp in length
*	24698	24777: gap of unknown length
*	24778	25699: contig of 922 bp in length
*	25700	25779: gap of unknown length
*	25780	26547: contig of 768 bp in length
*	26548	26627: gap of unknown length
*	26628	27682: contig of 1055 bp in length
*	27683	27762: gap of unknown length
*	27763	28171: contig of 409 bp in length
*	28172	28251: gap of unknown length
*	28252	28823: contig of 572 bp in length
*	28824	28903: gap of unknown length
*	28904	29965: contig of 1062 bp in length
*	29966	30045: gap of unknown length
*	30046	31173: contig of 1128 bp in length
*	31174	31253: gap of unknown length
*	31254	32156: contig of 903 bp in length
*	32157	32236: gap of unknown length
*	32237	32872: contig of 636 bp in length
*	32873	32952: gap of unknown length
*	32953	33817: contig of 865 bp in length
*	33818	33897: gap of unknown length
*	33898	35066: contig of 1169 bp in length
*	35067	35146: gap of unknown length
*	35147	37188: contig of 2042 bp in length
*	37189	37268: gap of unknown length
*	37269	38406: contig of 1138 bp in length
*	38407	38486: gap of unknown length
*	38487	40075: contig of 1589 bp in length
*	40076	40155: gap of unknown length
*	40156	41204: contig of 1049 bp in length
*	41205	41284: gap of unknown length
*	41285	42613: contig of 1329 bp in length
*	42614	42693: gap of unknown length
*	42694	45350: contig of 2657 bp in length
*	45351	45430: gap of unknown length

```
  *    45431    45914: contig of 484 bp in length
  *    45915    45994: gap of unknown length
  *    45995    46589: contig of 595 bp in length
  *    46590    46669: gap of unknown length
  *    46670    47222: contig of 553 bp in length
  *    47223    47302: gap of unknown length
  *    47303    47887: contig of 585 bp in length
  *    47888    47967: gap of unknown length
  *    47968    48553: contig of 586 bp in length
  *    48554    48633: gap of unknown length
  *    48634    49266: contig of 633 bp in length
  *    49267    49346: gap of unknown length
  *    49347    49944: contig of 598 bp in length
  *    49945    50024: gap of unknown length
  *    50025    50642: contig of 618 bp in length
  *    50643    50722: gap of unknown length
  *    50723    51227: contig of 505 bp in length
  *    51228    51307: gap of unknown length
  *    51308    52034: contig of 727 bp in length
  *    52035    52114: gap of unknown length
  *    52115    52686: contig of 572 bp in length
  *    52687    52766: gap of unknown length
  *    52767    53363: contig of 597 bp in length
  *    53364    53443: gap of unknown length
  *    53444    53993: contig of 550 bp in length
  *    53994    54073: gap of unknown length
  *    54074    54434: contig of 361 bp in length
  *    54435    54514: gap of unknown length
  *    54515    55129: contig of 615 bp in length
  *    55130    55209: gap of unknown length
  *    55210    55887: contig of 678 bp in length
  *    55888    55967: gap of unknown length
  *    55968    56510: contig of 543 bp in length
  *    56511    56590: gap of unknown length
  *    56591    57022: contig of 432 bp in length
  *    57023    57102: gap of unknown length
  *    57103    57394: contig of 292 bp in length
  *    57395    57474: gap of unknown length
  *    57475    58106: contig of 632 bp in length
  *    58107    58186: gap of unknown length
  *    58187    58920: contig of 734 bp in length
  *    58921    59000: gap of unknown length
  *    59001    59609: contig of 609 bp in length
  *    59610    59689: gap of unknown length
  *    59690    60371: contig of 682 bp in length
  *    60372    60451: gap of unknown length
  *    60452    61001: contig of 550 bp in length
  *    61002    61081: gap of unknown length
  *    61082    61656: contig of 575 bp in length
  *    61657    61736: gap of unknown length
  *    61737    62261: contig of 525 bp in length
  *    62262    62341: gap of unknown length
  *    62342    62865: contig of 524 bp in length
  *    62866    62945: gap of unknown length
  *    62946    63538: contig of 593 bp in length
  *    63539    63618: gap of unknown length
  *    63619    64136: contig of 518 bp in length
  *    64137    64216: gap of unknown length

  Query Match             39.7%;  Score 27;  DB 2;  Length 80866;
  Best Local Similarity   58.8%;  Pred. No. 17;
  Matches   30;  Conservative    6;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      15 TAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCG 65
           | ||| ||||||||||: : : |:|   |  ::|||| ||| | || |
Db  25511 TTGCATAAAGTTTTTACTGTTTACTAGTCATTTTGGAAGTGCGTTTGTTGG 25561


RESULT 12
AE001572_1/c
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AE001572 Accession AE001572
     Fragment Name      Begin     End
     AE001572_0             1    110000
     AE001572_1        100001    210000
     AE001572_2        200001    310000
     AE001572_3        300001    410000
     AE001572_4        400001    429825
Continuation (2 of 5) of AE001572 from base 100001 (AE001572 Drosophila melanogaster

  Query Match             39.7%;  Score 27;  DB 3;  Length 110000;
  Best Local Similarity   58.8%;  Pred. No. 17;
  Matches   30;  Conservative    6;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      15 TAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCG 65
           | ||| ||||||||||: : : |:|   |  ::|||| ||| | || |
Db  93289 TTGCATAAAGTTTTTACTGTTTACTAGTCATTTTGGAAGTGCGTTTGTTGG 93239


RESULT 13
AC095015
LOCUS       AC095015              190642 bp    DNA     linear   INV 15-SEP-2001
DEFINITION  Drosophila melanogaster, chromosome 3R, region 84A-84B, BAC clone
            BACR32J03, complete sequence.
ACCESSION   AC095015
VERSION     AC095015.1  GI:15624857
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 190642)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 3R, region 84A-84B
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190642)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
```

```
FEATURES             Location/Qualifiers
     source          1. .190642
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
                     /map="84A-84B"
                     /clone="BACR32J03 (D1350)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    55773 a  39290 c  38978 g  56601 t
ORIGIN

  Query Match             39.7%;  Score 27;  DB 3;  Length 190642;
  Best Local Similarity   58.8%;  Pred. No. 18;
  Matches   30;  Conservative    6;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      15 TAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCG 65
           | ||| |||||||||||: : : |:|   |  ::|||||  |||  | || |
Db  30503 TTGCATAAAGTTTTTACTGTTTACTAGTCATTTTGGAAGTGCGTTTGTTGG 30553


RESULT 14
AC107703
LOCUS       AC107703               208375 bp    DNA     linear   HTG 21-AUG-2002
DEFINITION  Mus musculus clone RP23-472H1, WORKING DRAFT SEQUENCE, 8 ordered
            pieces.
ACCESSION   AC107703
VERSION     AC107703.5  GI:22381033
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 208375)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-472H1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208375)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 208375)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 21, 2002 this sequence version replaced gi:21326325.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L18630
                Center clone name: 472_H_1
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 206755 bases at least Q40
                Consensus quality: 207261 bases at least Q30
                Consensus quality: 207461 bases at least Q20
                Insert size: 194000; agarose-fp
                Insert size: 207675; sum-of-contigs
                Quality coverage: 11.1 in Q20 bases; agarose-fp
                Quality coverage: 10.4 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *      1    53805: contig of 53805 bp in length
            *  53806 53905: gap of      100 bp
            *  53906    56169: contig of 2264 bp in length
            *  56170 56269: gap of      100 bp
            *  56270    62030: contig of 5761 bp in length
            *  62031 62130: gap of      100 bp
            *  62131    71647: contig of 9517 bp in length
            *  71648 71747: gap of      100 bp
            *  71748    84166: contig of 12419 bp in length
            *  84167 84266: gap of      100 bp
            *  84267   138240: contig of 53974 bp in length
            * 138241 138340: gap of      100 bp
            * 138341   186299: contig of 47959 bp in length
            * 186300 186399: gap of      100 bp
            * 186400   208375: contig of 21976 bp in length.
FEATURES             Location/Qualifiers
     source          1. .208375
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-472H1"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .53805
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    53906. .56169
```

```
                     /note="assembly_fragment"
     misc_feature    56270. .62030
                     /note="assembly_fragment"
     misc_feature    62131. .71647
                     /note="assembly_fragment"
     misc_feature    71748. .84166
                     /note="assembly_fragment"
     misc_feature    84267. .138240
                     /note="assembly_fragment"
     misc_feature    138341. .186299
                     /note="assembly_fragment"
     misc_feature    186400. .208375
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT     58490 a  45737 c  44986 g  58461 t    701 others
ORIGIN

  Query Match             39.7%;  Score 27;  DB 2;  Length 208375;
  Best Local Similarity   67.4%;  Pred. No. 18;
  Matches   29;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      21 AAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTT 63
           |||||| ||||:::| || | | ||| :| || ||| ||| ||
Db 159985 AAAGTTATTACTTTCAGCCAGGCAGGGTGAAAAGGCACGTATT 160027


RESULT 15
AE003673
LOCUS       AE003673               309357 bp    DNA     linear   INV 04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386040 section 3
            of 5, complete sequence.
ACCESSION   AE003673 AE002699
VERSION     AE003673.2  GI:10727114
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 309357)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
   PUBMED   10731132
REFERENCE   2  (bases 1 to 309357)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7298860.
FEATURES             Location/Qualifiers
     source          1. .309357
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
     gene            <13397. .>24361
                     /gene="Dfd"
                     /note="CG2189"
                     /map="84A5-84A5"
                     /db_xref="FLYBASE:FBan0002189"
                     /db_xref="FLYBASE:FBgn0000439"
     mRNA            join(13397. .14933,15319. .15419,15485. .15662,22801. .23187,
                     23445. .24361)
                     /gene="Dfd"
                     /product="CT7132"
                     /note="Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /db_xref="FLYBASE:FBan0002189"
                     /db_xref="FLYBASE:FBgn0000439"
     CDS             join(14147. .14933,15319. .15419,15485. .15662,22801. .23187,
                     23445. .23761)
                     /gene="Dfd"
                     /note="Dfd gene product; Nucleotide sequence of the Celera
                     sequence differs from the published sequence for this
                     transcript"
                     /codon_start=1
                     /protein_id="AAF54083.1"
                     /db_xref="GI:7298876"
                     /db_xref="FLYBASE:FBan0002189"
                     /db_xref="FLYBASE:FBgn0000439"
                     /translation="MSSFLMGYPHAPHHVQSPMSMGNGLDPKFPPLADDYHHYNGHYS
                     MTASTGHMSGAVGGGAGVGSVGGGGAGGMTGHPHSMHPADMVSDYMAHHHNPHSHSHS
                     HTHSLPHHHSNSAISGHHQASAGGYSSNYANATPPSHPHSHPHAHPHQSLGYYVHHAP
                     EFISAGAVHSDPTNGYGPAANVPNTSNGGGGGGSGAVLGGGAVGGSANGYYGGYGGGY
                     GTANGSVGSTHSQGHSPHSQMMDLPLQCSSTEPPTNTALGLQELGLKLEKRIEEAVPA
                     GQQLQELGMRLRCDDMGSENDDMSEEDRLMLDRSPDELGSNDNDDDLGDSDSDEDLMA
                     ETTDGERIIYPWMKKIHVAGVANGSYQPGMEPKRQRTAYTRHQILELEKEFHYNRYLT
                     RRRRIEIAHTLVLSERQIKIWFQNRRMKWKKDNKLPNTKNVRKKTVDANGNPTPVAKK
                     PTKRAASKKQQQAQQQQQSQQQQTQQTPVMNEVSECIRSDSLESIGDVSSSLGNPPYI
                     PAAPETTSSYPGSQQHLSNNNNNGSGNNNNNNNNNNSNLNNNNNNNQMGHTNLHGHLQ
                     QQQSDLMTNLQLHIKQDYDLTAL"
     gene            complement(<46894. .>70354)
                     /gene="Scr"
                     /note="CG1030"
                     /map="84A5-84B1"
                     /db_xref="FLYBASE:FBan0001030"
                     /db_xref="FLYBASE:FBgn0003339"
     mRNA            complement(join(46894. .47918,63464. .64443,70145. .70354))
                     /gene="Scr"
                     /product="CT1096"
                     /note="Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /db_xref="FLYBASE:FBan0001030"
                     /db_xref="FLYBASE:FBgn0003339"
     CDS             complement(join(47611. .47918,63464. .64409))
                     /gene="Scr"
                     /note="Scr gene product; Nucleotide sequence of the Celera
```

```
                sequence differs from the published sequence for this
                transcript"
                /codon_start=1
                /protein_id="AAF54082.1"
                /db_xref="GI:7298875"
                /db_xref="FLYBASE:FBan0001030"
                /db_xref="FLYBASE:FBgn0003339"
                /translation="MDPDCFAMSSYQFVNSLASCYPQQMNPQQNHPGAGNSSAGGSGG
                GAGGSGGVVPSGGTNGGQGSAGAATPGANDYFPAAAAYTPNLYPNTPQAHYANQAAYG
                GQGNPDMVDYTQLQPQRLLLQQQQQQQQQQHAHAAAAVAAQQQQQLAQQQHPQQQQQQ
                QQANISCKYANDPVTPGGSGGGGVSGSNNNNNSANSNNNNSQSLASPQDLSTRDISPK
                LSPSSVVESVARSLNKGVLGGSLAAAAAAAGLNNNHSGSGVSGGPGNVNVPMHSPGGG
                DSDSESDSGNEAGSSQNSGNGKKNPPQIYPWMKRVHLGTSTVNANGETKRQRTSYTRY
                QTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEHKMASMNIVP
                YHMGPYGHPYHQFDIHPSQFAHLSA"
gene            <86526. .>88310
                /gene="ftz"
                /note="CG2047"
                /map="84B1-84B1"
                /db_xref="FLYBASE:FBan0002047"
                /db_xref="FLYBASE:FBgn0001077"
mRNA            join(86526. .86589,86623. .87370,87518. .88310)
                /gene="ftz"
                /product="CT6611"
                /note="Nucleotide sequence of the Celera sequence differs
                from the published sequence for this transcript."
                /db_xref="FLYBASE:FBan0002047"
                /db_xref="FLYBASE:FBgn0001077"
CDS             join(86623. .87370,87518. .88002)
                /gene="ftz"
                /note="ftz gene product; Nucleotide sequence of the Celera
                sequence differs from the published sequence for this
                transcript"
                /codon_start=1
                /protein_id="AAF54081.1"
                /db_xref="GI:7298874"
                /db_xref="FLYBASE:FBan0002047"
                /db_xref="FLYBASE:FBgn0001077"
                /translation="MATTNSQSHYSYADNMNMYNMYHPHSLPPTYYDNSGSNAYYQNT
                SNYQGYYPQESYSESCYYYNNQEQVTTQTVPPVQPTTPPPKATKRKAEDDAASIIAAV
                EERPSTLRALLTNPVKKLKYTPDYFYTTVEQVKKAPAVSTKVTASPAPSYDQEYVTVP
                TPSASEDVDYLDVYSPQSQTQKLKNGDFATPPPTTPTSLPPLEGISTPPQSPGEKSSS
                AVSQEINHRIVTAPNGAGDFNWSHIEETLASDCKDSKRTRQTYTRYQTLELEKEFHFN
                RYITRRRRIDIANALSLSERQIKIWFQNRRMKSKKDRTLDSSPEHCGAGYTAMLPPLE
                ATSTATTGAPSVPVPMYHHHQTTAAYPAYSHSHSHGYGLLNDYPQQQTHQQYDAYPQQ
                YQHQCSYQQHPQDLYHLS"
gene            complement(<118482. .>223587)
                /gene="Antp"
                /note="CG1028"
                /map="84B1-84B4"
                /db_xref="FLYBASE:FBan0001028"
                /db_xref="FLYBASE:FBgn0000095"
mRNA            complement(join(118482. .120767,130742. .130967,
                131102. .131140,131519. .132295,193615. .193704,
                221238. .221367,221470. .222034,222070. .222260,
                222819. .223415,223431. .223587))
                /gene="Antp"
                /product="CT30093"
                /db_xref="FLYBASE:FBan0001028"
                /db_xref="FLYBASE:FBgn0000095"
CDS             complement(join(120517. .120767,130742. .130967,
                131102. .131140,131519. .132139))
                /gene="Antp"
                /note="Antp gene product"
                /codon_start=1
                /protein_id="AAG22205.1"
                /db_xref="GI:10727115"
                /db_xref="FLYBASE:FBan0001028"
                /db_xref="FLYBASE:FBgn0000095"
                /translation="MTMSTNNCESMTSYFTNSYMGADMHHGHYPGNGVTDLDAQQMHH
                YSQNANHQGNMPYPRFPPYDRMPYYNGQGMDQQQQHQVYSRPDSPSSQVGGVMPQAQT
                NGQLGVPQQQQQQQQQPSQNQQQQAQQAPQQLQQQLPQVTQQVTHPQQQQQQPVVYA
                SCKLQAAVGGLGMVPEGGSPPLVDQMSGHHMNAQMTLPHHMGHPQAQLGYTDVGVPDV
                TEVHQNHHNMGMYQQQSGVPPVGAPPQGMMHQGQGPPQMHQGHPGQHTPPSQNPNSQS
                SGMPSPLYPWMRSQFGKCQERKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHAL
                CLTERQIKIWFQNRRMKWKKENKTKGEPGSGGEGDEITPPNSPQ"
gene            <275368. .>277531
                /gene="Sodh-1"
                /note="CG1982"
                /map="84C1-84C1"
                /db_xref="FLYBASE:FBan0001982"
                /db_xref="FLYBASE:FBgn0024289"
mRNA            join(275368. .275682,275741. .275774,276161. .276630,
```

Query Match 39.7%; Score 27; DB 3; Length 309357;
Best Local Similarity 58.8%; Pred. No. 19;
Matches 30; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

```
Qy      15 TAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCG 65
           | ||| |||||||||||: : : |:|   |  ::|||||| |||  | || |
Db  85770 TTGCATAAAGTTTTTACTGTTTACTAGTCATTTTGGAAGTGCGTTTGTTGG 85820
```

Search completed: March 17, 2003, 11:24:06
Job time : 920.892 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:42:12 ; Search time 389.72 Seconds
(without alignments)
392.938 Million cell updates/sec

Title: US-09-836-439-1
Perfect score: 68
Sequence: 1 ccttccaacctacgtagcag.....ggaagggcgcgttttcgcgc 68

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	28	41.2	68	18	AAV09396	Chimeric repair ve
	2	28	41.2	68	19	AAV12904	Chimeric mutationa
	3	28	41.2	68	20	AAX19657	Oligonucleotide SE
	4	27	39.7	68	19	AAV52984	Oligonucleotide us
c	5	27	39.7	3785	23	ABL03968	Drosophila melanog
	6	26.4	38.8	68	18	AAV09395	Chimeric repair ve
	7	26.4	38.8	68	19	AAV12903	Chimeric mutationa
	8	26.4	38.8	68	20	AAX19656	Oligonucleotide SE
	9	25.6	37.6	533	24	ABQ51034	Oligonucleotide fo
c	10	25.6	37.6	533	24	ABQ51035	Oligonucleotide fo
	11	25.4	37.4	435	24	ABL38127	Human colon tumour
c	12	25.4	37.4	1280	22	AAI99525	Human polynucleoti
	13	25.4	37.4	2257	24	AAL44063	Human genset metab
	14	25.4	37.4	2454	22	AAK52451	Human polynucleoti
	15	25.4	37.4	2454	22	AAK52452	Human polynucleoti
	16	25.4	37.4	2643	22	AAK51468	Human polynucleoti
	17	25.4	37.4	2727	22	AAD16945	Novel human protei
	18	25.4	37.4	2730	22	AAD16951	Novel human protei
	19	25.4	37.4	2874	22	AAD16947	Novel human protei
	20	25.4	37.4	2877	22	AAD16953	Novel human protei
	21	25.4	37.4	2913	22	AAK51467	Human polynucleoti
	22	25.4	37.4	2979	22	AAD16949	Novel human protei
	23	25.4	37.4	2982	22	AAD16955	Novel human protei
	24	25.4	37.4	3035	24	ABA97973	Human pancreatic c
	25	25.4	37.4	3129	22	AAD16944	Novel human protei
	26	25.4	37.4	3132	22	AAD16950	Novel human protei
	27	25.4	37.4	3275	20	AAZ07192	Human lung tumour
	28	25.4	37.4	3275	21	AAC79145	Human lung tumour-
	29	25.4	37.4	3275	23	AAD23220	Human lung tumour-
	30	25.4	37.4	3276	22	AAD16946	Novel human protei
	31	25.4	37.4	3279	22	AAD16952	Novel human protei
	32	25.4	37.4	3381	22	AAD16948	Novel human protei
	33	25.4	37.4	3384	22	AAD16954	Novel human protei
	34	25.4	37.4	4342	22	AAD16962	Novel human protei
	35	25.2	37.1	4501	24	ABK33939	Human DNA for stag
	36	25.2	37.1	8056	24	ABK34030	Human DNA for stag
	37	25.2	37.1	10996	22	AAS46806	Tumour suppressor
	38	25.2	37.1	10996	24	ABK28466	DNA transcription
	39	24.8	36.5	68	18	AAV09393	Chimeric repair ve
	40	24.8	36.5	68	18	AAV09394	Chimeric repair ve
	41	24.8	36.5	68	19	AAV12901	Chimeric mutationa
	42	24.8	36.5	68	19	AAV12902	Chimeric mutationa
	43	24.8	36.5	68	20	AAX19655	Oligonucleotide SE
	44	24.8	36.5	68	20	AAX19654	Oligonucleotide SE
	45	24.8	36.5	68	22	AAF57158	Chimeric DNA/RNA o

ALIGNMENTS

RESULT 1
AAV09396
ID AAV09396 standard; DNA; 68 BP.
XX
AC AAV09396;
XX
DT 14-MAY-1998 (first entry)
XX
DE Chimeric repair vector (CRV) SC4.
XX
KW Chimeric repair vector; CRV; treatment; genetic mutation; repair;
KW sickle cell disease; beta-thalassemia; Gaucher disease; emphysema;
KW hypercholesterolaemia; haemophilia; DNA/RNA hybrid; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_feature 1..29
FT /*tag= a
FT /note= "DNA nucleotides"
FT misc_RNA 30..39
FT /*tag= b
FT /note= "RNA nucleotides"
FT misc_feature 40..44
FT /*tag= c
FT /note= "DNA nucleotides"
FT misc_RNA 45..54
FT /*tag= d
FT /note= "RNA nucleotides"
FT misc_feature 55..68
FT /*tag= e

```
FT                   /note= "DNA nucleotides"
XX
PN   WO9741141-A1.
XX
PD   06-NOV-1997.
XX
PF   01-MAY-1997;   97WO-US07362.
XX
PR   01-MAY-1996;   96US-0640517.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Cole-strauss A,  Kmiec EB,  Yoon K;
XX
DR   WPI; 1997-549675/50.
XX
PT   Chimeric nucleic acid repair vectors - used for treating diseases
PT   such as sickle cell disease, beta-thalassemia, Gaucher disease,
PT   hyper:cholesterolaemia, emphysema or haemophilia
XX
PS   Disclosure; Fig 3; 79pp; English.
XX
CC   This chimeric repair vector (CRV) SC4 is used in comparative studies on
CC   the experimental use of a CRV SC1 designed to repair the mutation found
CC   in Sickle cell disease beta-globin and the beta-globin of a HSC. The CRV
CC   designed to repair the mutation contains a nucleic acid having at most
CC   one 3' end and one 5' end comprising a segment of unpaired bases
CC   disposed. The unpaired bases separate the nucleic acid into a first
CC   strand and a second strand, comprising a first region and a second region
CC   respectively, each region having at least 15 nucleotides. Each nucleotide
CC   of the first region is Watson-Crick paired to a nucleotide of the second
CC   region and the first region comprises at least 8 ribonucleotides, which
CC   are Watson-Crick paired to 2'-deoxynucleotides, which ribonucleotides
CC   form at least one ribonucleotide segment of at least 3 ribonucleotides
CC   and the sequence of the first or the second region is the sequence of a
CC   fragment of a wild-type allele of a human gene. The CRVs can be used for
CC   repairing genetic mutations in cells for re-introducing into a patient
CC   for treating diseases. They can be used for treating, sickle cell
CC   disease, beta-thalassemia, familial hypercholesterolaemia, Gaucher
CC   disease, emphysema or haemophilia.
XX
SQ   Sequence 68 BP; 11 A; 19 C; 19 G; 13 T; 6 U; 0 other;

  Query Match             41.2%;  Score 28;  DB 18;  Length 68;
  Best Local Similarity   66.7%;  Pred. No. 0.57;
  Matches   40;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      9 CCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCGCGC 68
          |||  | || |||  | |||| |  ||| || |  |||    | | ||||||||||||||
Db      9 CCTGAGGAGAAGACTGCTTTTGCAGUCUUCUCCTCAGGAGUCAGGUGCGCGTTTTCGCGC 68


RESULT 2
AAV12904
ID   AAV12904 standard; DNA; 68 BP.
XX
AC   AAV12904;
XX
DT   17-JUN-1998  (first entry)
XX
DE   Chimeric mutational vector SC4.
XX
KW   Chimeric mutational vector; alkaline phosphatase gene; gene repair;
KW   disease-related mutation; human; Gaucher's disease; sickle-cell anaemia;
KW   thalassemia; familial hypercholesterolaemia; emphysema; therapy;
KW   circular; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..25
FT                   /*tag=  a
FT                   /note= "binds to nucleotides 54 to 30"
FT   misc_feature    30..54
FT                   /*tag=  b
FT                   /note= "binds to nucleotides 25 to 1"
FT   misc_RNA        30..39
FT                   /*tag=  c
FT   misc_RNA        45..54
FT                   /*tag=  d
FT   misc_feature    55..59
FT                   /*tag=  e
FT                   /note= "binds to nucleotides 68 to 64"
FT   misc_feature    64..68
FT                   /*tag=  f
FT                   /note= "binds to nucleotides 59 to 55"
XX
PN   WO9748714-A1.
XX
PD   24-DEC-1997.
XX
PF   16-JUN-1997;   97WO-US10538.
XX
PR   17-JUN-1996;   96US-0664487.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Kmiec EB;
XX
DR   WPI; 1998-063068/06.
XX
PT   Oligonucleotide for altering a genomic sequence in eukaryotes -
PT   particularly for correcting disease-related mutation(s) and for
PT   production of transgenic animals and plants
XX
PS   Example 7.2; Fig 3; 68pp; English.
XX
CC   This sequence represents a nucleotide analogue of the invention, termed a
CC   chimeric mutational vector (CMV). This sequence is directed against the
CC   beta-globin gene. The CMVs (I) are for altering a gene in a
CC   eukaryotic cell, and comprise: (a) a first strand (S1) having at least 15
CC   nucleotides (nt); at least 3 nuclease-resistant ribo-type nt (nt') and at
CC   least 3 contiguous ribo-type nt, the same as, or additional to, the nt';
CC   and (b) a second strand (S2) in which the nt are Watson-Crick (WC) paired
CC   to the nt in S1. The contiguous ribo-type nt in S1 are WC-paired to
CC   2'-deoxyribo-type nt, and at least one ribo-type nt is other than a
CC   2'-O-methyl substituted nt. nt' are 2'-AX-nucleosides; 2'-AX-nucleosides
CC   or 2'-AR-nucleotides; A = oxygen, fluoro, chloro or bromo; when A = O,
CC   X = hydrogen or 1-6C alkyl and R = 1-6C alkyl; when A is halo then R and
CC   X are absent. (I) are used to repair a disease-related mutation in human
CC   cells (e.g. those associated with Gaucher's disease, sickle-cell anaemia,
CC   thalassemia, familial hypercholesterolaemia, emphysema etc.). The
CC   chimeric mutation vectors are also used to inactivate specific genes,
CC   i.e. to generate transgenic ('knockout') animals or plants. They are also
CC   used for biomedical research and for pharmaceutical production. Any
CC   eukaryotic gene of known sequence can be altered, by replacement,
CC   deletion or addition.
XX
SQ   Sequence 68 BP; 11 A; 19 C; 19 G; 13 T; 6 U; 0 other;

  Query Match             41.2%;  Score 28;  DB 19;  Length 68;
  Best Local Similarity   66.7%;  Pred. No. 0.57;
  Matches   40;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      9 CCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCGCGC 68
          |||  | || |||  | |||| |  ||| || |  |||    | | ||||||||||||||
Db      9 CCTGAGGAGAAGACTGCTTTTGCAGUCUUCUCCTCAGGAGUCAGGUGCGCGTTTTCGCGC 68


RESULT 3
AAX19657
ID   AAX19657 standard; DNA; 68 BP.
XX
AC   AAX19657;
XX
DT   02-JUN-1999  (first entry)
```

```
XX
DE   Oligonucleotide SEQ ID NO:96.
XX
KW   Genome; genetic lesion; haematopoietic stem cell; hepatocyte; RNase;
KW   human wild-type allele; mutation; sickle cell anaemia; thalassemia;
KW   Gaucher's disease; glucocerebrosidase gene; hypercholesterolaemia;
KW   emphysema; haemophilia; Christmas disease; ss.
XX
OS   Synthetic.
XX
PN   US5888983-A.
XX
PD   30-MAR-1999.
XX
PF   05-AUG-1997;   97US-0906265.
XX
PR   05-AUG-1997;   97US-0906265.
PR   01-MAY-1996;   96US-0640517.
PR   17-JUN-1996;   96US-0664487.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Cole-Strauss AD,  Kmiec EB;
XX
DR   WPI; 1999-243264/20.
XX
PT   Double-stranded oligonucleotides with containing a human wild-type
PT   allele  - useful for repairing mutations in human cells,
PT   particularly those causing sickle cell anaemia or thalassemia
XX
PS   Disclosure; Column 61-62; 40pp; English.
XX
CC   The present invention describes double-stranded oligonucleotides (I)
CC   containing fragments of wild-type human alleles. (I) are used to repair
CC   disease associated mutations in human cells. (I) are preferably used to
CC   treat sickle cell anaemia or thalassemia (mutations in the beta-globin
CC   gene, including the promoter region), or Gaucher's disease (mutations in
CC   the glucocerebrosidase gene), in haematopoietic cells. (I) may also be
CC   used to treat familial hypercholesterolaemia (mutations in the low-
CC   density lipoprotein receptor gene), emphysema (the alpha 1-anti-trypsin
CC   gene), haemophilia (the factor VIII gene) or Christmas disease (the
CC   factor IX gene), in hepatocytic cells. (I) provides repair of small
CC   genetic mutations. The present sequence represents an oligonucleotide
CC   from the present invention.
XX
SQ   Sequence 68 BP; 11 A; 19 C; 19 G; 13 T; 6 U; 0 other;

  Query Match             41.2%;  Score 28;  DB 20;  Length 68;
  Best Local Similarity   66.7%;  Pred. No. 0.57;
  Matches   40;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       9 CCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCGCGC 68
           |||  | || |||  | |||| |  ||| || |  |||    | | ||||||||||||||
Db       9 CCTGAGGAGAAGACTGCTTTTGCAGUCUUCUCCTCAGGAGUCAGGUGCGCGTTTTCGCGC 68


RESULT 4
AAV52984
ID   AAV52984 standard; DNA; 68 BP.
XX
AC   AAV52984;
XX
DT   04-DEC-1998  (first entry)
XX
DE   Oligonucleotide used in the course of the invention.
XX
KW   Chimeric oligonucleotide; specific alteration; target sequence;
KW   intramolecular duplex stability; intermolecular duplex stability;
KW   nuclease degradation; chemical stability; hydrolysis resistance;
KW   degradation resisitance; A-type helix formation; stable conformation;
KW   DNA/RNA hybrid; cyclic; circular; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    26..29
FT                   /*tag=  a
FT                   /note= "nucleotides replacing pentaethylene oxide (PEO)
FT                          loop portion"
XX
PN   WO9839353-A1.
XX
PD   11-SEP-1998.
XX
PF   19-FEB-1998;   98WO-US03224.
XX
PR   03-MAR-1997;   97US-0039244.
XX
PA   (PEKE ) PERKIN-ELMER CORP.
XX
PI   Andrus A,  Kuimelis RG;
XX
DR   WPI; 1998-495787/42.
XX
PT   New chimeric oligo:nucleotide(s) - which are useful for causing
PT   specific alterations in target nucleic acids
XX
PS   Example 3; Page 33; 56pp; English.
XX
CC   This oligonucleotide is used in the course of the invention. The
CC   invention provides chimeric oligonucleotides comprising a stem portion
CC   having one or more nucleotides selected from 2'-O-alkyl-ribonucleotides,
CC   2'-O-allyl-ribonucleotides, 2'-allyl-ribonucleotides, 2'-halo-
CC   ribonucleotides, 2'-O-methoxyethyl-ribonucleotides, 2'-branching group-
CC   ribonucleotides or 2'-O-branching group-ribonucleotides, and one or more
CC   loop portions. The chimeric oligonucleotides may be used for introduction
CC   of specific alterations into target nucleic acid sequences located in
CC   living organisms. They have increased intramolecular and intermolecular
CC   duplex stability, increased resistance to nuclease degradation and
CC   increased chemical stability, resistance to hydrolysis and degradation,
CC   as compared to known chimeric oligonucleotides. They have structures
CC   which induce intramolecular and intermolecular A-type helix formation.
CC   They have highly stable conformations, as measured by thermal melting
CC   measurements.
XX
SQ   Sequence 68 BP; 13 A; 17 C; 17 G; 21 T; 0 other;

  Query Match             39.7%;  Score 27;  DB 19;  Length 68;
  Best Local Similarity   58.2%;  Pred. No. 1.3;
  Matches   39;  Conservative    3;  Mismatches   25;  Indels    0;  Gaps    0;

Qy       2 CTTCCAACCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTT 61
           |||||||||    |  ||   |   ||||   :   ||  |    |::|||||| |||||||
Db       2 CTTCCAACGAGGGATGCTCTACTATTTTTAGTAGAGCATCCCTCGTTGGAAGTGCGCGTT 61

Qy      62 TTCGCGC 68
           |||||||
Db      62 TTCGCGC 68


RESULT 5
ABL03968/c
ID   ABL03968 standard; cDNA; 3785 BP.
XX
AC   ABL03968;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 6386.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
```

```
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB59865.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 6386; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3785 BP; 876 A; 851 C; 982 G; 1076 T; 0 other;

  Query Match             39.7%;  Score 27;  DB 23;  Length 3785;
  Best Local Similarity   58.8%;  Pred. No. 3.3;
  Matches   30;  Conservative    6;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      15 TAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCG 65
           | ||| |||||||||||: : : |:|   |  ::||||| |||  | || |
Db    3541 TTGCATAAAGTTTTTACTGTTTACTAGTCATTTTGGAAGTGCGTTTGTTGG 3491


RESULT 6
AAV09395
ID   AAV09395 standard; DNA; 68 BP.
XX
AC   AAV09395;
XX
DT   14-MAY-1998  (first entry)
XX
DE   Chimeric repair vector (CRV) SC3.
XX
KW   Chimeric repair vector; CRV; treatment; genetic mutation; repair;
KW   sickle cell disease; beta-thalassemia; Gaucher disease; emphysema;
KW   hypercholesterolaemia; haemophilia; DNA/RNA hybrid; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..29
FT                   /*tag=  a
FT                   /note= "DNA nucleotides"
FT   misc_RNA        30..39
FT                   /*tag=  b
FT                   /note= "RNA nucleotides"
FT   misc_feature    40..44
FT                   /*tag=  c
FT                   /note= "DNA nucleotides"
FT   misc_RNA        45..54
FT                   /*tag=  d
FT                   /note= "RNA nucleotides"
FT   misc_feature    55..68
FT                   /*tag=  e
FT                   /note= "DNA nucleotides"
XX
PN   WO9741141-A1.
XX
PD   06-NOV-1997.
XX
PF   01-MAY-1997;   97WO-US07362.
XX
PR   01-MAY-1996;   96US-0640517.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Cole-strauss A,  Kmiec EB,  Yoon K;
XX
DR   WPI; 1997-549675/50.
XX
PT   Chimeric nucleic acid repair vectors - used for treating diseases
PT   such as sickle cell disease, beta-thalassemia, Gaucher disease,
PT   hyper:cholesterolaemia, emphysema or haemophilia
XX
PS   Disclosure; Fig 3; 79pp; English.
XX
CC   This chimeric repair vector (CRV) SC3 is used in comparative studies on
CC   the experimental use of a CRV SC1 designed to repair the mutation found
CC   in Sickle cell disease beta-globin and the beta-globin of a HSC. The CRV
CC   designed to repair the mutation contains a nucleic acid having at most
CC   one 3' end and one 5' end comprising a segment of unpaired bases
CC   disposed. The unpaired bases separate the nucleic acid into a first
CC   strand and a second strand, comprising a first region and a second region
CC   respectively, each region having at least 15 nucleotides. Each nucleotide
CC   of the first region is Watson-Crick paired to a nucleotide of the second
CC   region and the first region comprises at least 8 ribonucleotides, which
CC   are Watson-Crick paired to 2'-deoxynucleotides, which ribonucleotides
CC   form at least one ribonucleotide segment of at least 3 ribonucleotides
CC   and the sequence of the first or the second region is the sequence of a
CC   fragment of a wild-type allele of a human gene. The CRVs can be used for
CC   repairing genetic mutations in cells for re-introducing into a patient
CC   for treating diseases. They can be used for treating, sickle cell
CC   disease, beta-thalassemia, familial hypercholesterolaemia, Gaucher
CC   disease, emphysema or haemophilia.
XX
SQ   Sequence 68 BP; 12 A; 18 C; 18 G; 14 T; 6 U; 0 other;

  Query Match             38.8%;  Score 26.4;  DB 18;  Length 68;
  Best Local Similarity   65.0%;  Pred. No. 2.2;
  Matches   39;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       9 CCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCGCGC 68
           |||  | || |||  | |||| |  ||| || |  |||    |   ||||||||||||||
Db       9 CCTGAGGAGAAGACTGCTTTTGCAGUCUUCUCCTCAGGAGUCAGAUGCGCGTTTTCGCGC 68


RESULT 7
AAV12903
ID   AAV12903 standard; DNA; 68 BP.
XX
AC   AAV12903;
XX
DT   17-JUN-1998  (first entry)
XX
DE   Chimeric mutational vector SC3.
XX
KW   Chimeric mutational vector; alkaline phosphatase gene; gene repair;
KW   disease-related mutation; human; Gaucher's disease; sickle-cell anaemia;
KW   thalassemia; familial hypercholesterolaemia; emphysema; therapy;
KW   circular; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
```

```
FT   misc_feature    1..25
FT                   /*tag= a
FT                   /note= "binds to nucleotides 54 to 30"
FT   misc_feature    30..54
FT                   /*tag= b
FT                   /note= "binds to nucleotides 25 to 1"
FT   misc_RNA        30..39
FT                   /*tag= c
FT   misc_RNA        45..54
FT                   /*tag= d
FT   misc_feature    55..59
FT                   /*tag= e
FT                   /note= "binds to nucleotides 68 to 64"
FT   misc_feature    64..68
FT                   /*tag= f
FT                   /note= "binds to nucleotides 59 to 55"
XX
PN   WO9748714-A1.
XX
PD   24-DEC-1997.
XX
PF   16-JUN-1997;   97WO-US10538.
XX
PR   17-JUN-1996;   96US-0664487.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Kmiec EB;
XX
DR   WPI; 1998-063068/06.
XX
PT   Oligonucleotide for altering a genomic sequence in eukaryotes -
PT   particularly for correcting disease-related mutation(s) and for
PT   production of transgenic animals and plants
XX
PS   Example 7.2; Fig 3; 68pp; English.
XX
CC   This sequence represents a nucleotide analogue of the invention, termed a
CC   chimeric mutational vector (CMV). This sequence is directed against the
CC   beta-globin gene. The CMVs (I) are for altering a gene in a
CC   eukaryotic cell, and comprise: (a) a first strand (S1) having at least 15
CC   nucleotides (nt); at least 3 nuclease-resistant ribo-type nt (nt') and at
CC   least 3 contiguous ribo-type nt, the same as, or additional to, the nt';
CC   and (b) a second strand (S2) in which the nt are Watson-Crick (WC) paired
CC   to the nt in S1. The contiguous ribo-type nt in S1 are WC-paired to
CC   2'-deoxyribo-type nt, and at least one ribo-type nt is other than a
CC   2'-O-methyl substituted nt. nt' are 2'-AX-nucleosides; 2'-AX-nucleosides
CC   or 2'-AR-nucleotides; A = oxygen, fluoro, chloro or bromo; when A = O,
CC   X = hydrogen or 1-6C alkyl and R = 1-6C alkyl; when A is halo then R and
CC   X are absent. (I) are used to repair a disease-related mutation in human
CC   cells (e.g. those associated with Gaucher's disease, sickle-cell anaemia,
CC   thalassemia, familial hypercholesterolaemia, emphysema etc.). The
CC   chimeric mutation vectors are also used to inactivate specific genes,
CC   i.e. to generate transgenic ('knockout') animals or plants. They are also
CC   used for biomedical research and for pharmaceutical production. Any
CC   eukaryotic gene of known sequence can be altered, by replacement,
CC   deletion or addition.
XX
SQ   Sequence 68 BP; 12 A; 18 C; 18 G; 14 T; 6 U; 0 other;

  Query Match             38.8%;  Score 26.4;  DB 19;  Length 68;
  Best Local Similarity   65.0%;  Pred. No. 2.2;
  Matches   39;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       9 CCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCGCGC 68
           |||  | || |||  | |||| |  ||| || |  |||    |   |||||||||||||
Db       9 CCTGAGGAGAAGACTGCTTTTGCAGUCUUCUCCTCAGGAGUCAGAUGCGCGTTTTCGCGC 68


RESULT 8
AAX19656
ID   AAX19656 standard; DNA; 68 BP.
XX
AC   AAX19656;
XX
DT   02-JUN-1999  (first entry)
XX
DE   Oligonucleotide SEQ ID NO:95.
XX
KW   Genome; genetic lesion; haematopoietic stem cell; hepatocyte; RNase;
KW   human wild-type allele; mutation; sickle cell anaemia; thalassemia;
KW   Gaucher's disease; glucocerebrosidase gene; hypercholesterolaemia;
KW   emphysema; haemophilia; Christmas disease; ss.
XX
OS   Synthetic.
XX
PN   US5888983-A.
XX
PD   30-MAR-1999.
XX
PF   05-AUG-1997;   97US-0906265.
XX
PR   05-AUG-1997;   97US-0906265.
PR   01-MAY-1996;   96US-0640517.
PR   17-JUN-1996;   96US-0664487.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Cole-Strauss AD,  Kmiec EB;
XX
DR   WPI; 1999-243264/20.
XX
PT   Double-stranded oligonucleotides with containing a human wild-type
PT   allele  - useful for repairing mutations in human cells,
PT   particularly those causing sickle cell anaemia or thalassemia
XX
PS   Disclosure; Column 59-60; 40pp; English.
XX
CC   The present invention describes double-stranded oligonucleotides (I)
CC   containing fragments of wild-type human alleles. (I) are used to repair
CC   disease associated mutations in human cells. (I) are preferably used to
CC   treat sickle cell anaemia or thalassemia (mutations in the beta-globin
CC   gene, including the promoter region), or Gaucher's disease (mutations in
CC   the glucocerebrosidase gene), in haematopoietic cells. (I) may also be
CC   used to treat familial hypercholesterolaemia (mutations in the low-
CC   density lipoprotein receptor gene), emphysema (the alpha 1-anti-trypsin
CC   gene), haemophilia (the factor VIII gene) or Christmas disease (the
CC   factor IX gene), in hepatocytic cells. (I) provides repair of small
CC   genetic mutations. The present sequence represents an oligonucleotide
CC   from the present invention.
XX
SQ   Sequence 68 BP; 12 A; 18 C; 18 G; 14 T; 6 U; 0 other;

  Query Match             38.8%;  Score 26.4;  DB 20;  Length 68;
  Best Local Similarity   65.0%;  Pred. No. 2.2;
  Matches   39;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       9 CCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCGCGC 68
           |||  | || |||  | |||| |  ||| || |  |||    |   |||||||||||||
Db       9 CCTGAGGAGAAGACTGCTTTTGCAGUCUUCUCCTCAGGAGUCAGAUGCGCGTTTTCGCGC 68


RESULT 9
ABQ51034
ID   ABQ51034 standard; DNA; 533 BP.
XX
AC   ABQ51034;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 37625.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
```

XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 533 BP; 89 A; 62 C; 208 G; 174 T; 0 other;

Query Match 37.6%; Score 25.6; DB 24; Length 533;
Best Local Similarity 55.4%; Pred. No. 7.1;
Matches 31; Conservative 6; Mismatches 19; Indels 0; Gaps 0;

Qy 12 ACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCGCG 67
| || || | | ||||| ::: :| | |||::||||| ||| | ||| | |
Db 1 ATGTCGCGGGATTTTTTTTATTTTATGTAGGAAGGTTGGAAGCGCGAGATTTAGGG 56

RESULT 10
ABQ51035/c
ID ABQ51035 standard; DNA; 533 BP.
XX
AC ABQ51035;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 37626.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 533 BP; 174 A; 208 C; 62 G; 89 T; 0 other;

Query Match 37.6%; Score 25.6; DB 24; Length 533;
Best Local Similarity 55.4%; Pred. No. 7.1;
Matches 31; Conservative 6; Mismatches 19; Indels 0; Gaps 0;

Qy 12 ACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCGCG 67
| || || | | ||||| ::: :| | |||::||||| ||| | ||| | |
Db 533 ATGTCGCGGGATTTTTTTTATTTTATGTAGGAAGGTTGGAAGCGCGAGATTTAGGG 478

RESULT 11
ABL38127
ID ABL38127 standard; cDNA; 435 BP.
XX
AC ABL38127;
XX
DT 08-APR-2002 (first entry)
XX
DE Human colon tumour antigen polynucleotide SEQ ID NO:1716.
XX
KW Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
KW colon tumour metastatic antigen; diagnosis; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200196388-A2.
XX
PD 20-DEC-2001.
XX
PF 08-JUN-2001; 2001WO-US18557.

XX
PR 09-JUN-2000; 2000US-210899P.
PR 20-FEB-2001; 2001US-270216P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Jiang Y, Harlocker SL, Secrist H;
XX
DR WPI; 2002-114514/15.
XX
PT Novel isolated colon tumor polynucleotide differentially expressed in
PT colon tumor or colon metastatic tumor and polypeptides encoded by them,
PT useful for inhibiting development of cancer in patient -
XX
PS Claim 1; SEQ ID 1716; 105pp; English.
XX
CC ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (I)
CC which were isolated from human colon tumour and colon metastatic tumour
CC cDNA libraries. (I) have cytostatic activity and can be used in vaccine
CC production. (I) can be used for stimulating and/or expanding T cells
CC specific for a tumour protein on contact with the T cells. They are also
CC useful for inhibiting the development of cancer in a patient. (I) can be
CC used as probes or primers for nucleic acid hybridisation, for preparing
CC mutant species primers, or primers for use in genetic constructions. (I)
CC can be used in the diagnosis of a colon tumour.
XX
SQ Sequence 435 BP; 114 A; 98 C; 91 G; 130 T; 2 other;

Query Match 37.4%; Score 25.4; DB 24; Length 435;
Best Local Similarity 55.9%; Pred. No. 8;
Matches 33; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

```
Qy       2 CTTCCAACCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGT 60
           ||||  | | || ||| ||||| |||| |  ::| ||:  |||| ::  |  | |  ||
Db     104 CTTCTGAACAACTTAGGAGAAACTTTTGATCTTCAGCTTGGTAGATTTAATTGCCCAGT 162
```

RESULT 12
AAI99525/c
ID AAI99525 standard; cDNA; 1280 BP.
XX
AC AAI99525;
XX
DT 07-JAN-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 23.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX
OS Homo sapiens.
XX
PN WO200155173-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01356.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451924/48.
DR P-PSDB; AAM99927.
XX
PT New nucleic acids and polypeptides, useful for treating, preventing or
PT ameliorating human disorders and diseases -
XX
PS Claim 1; SEQ ID NO 23; 465pp + Sequence Listing; English.
XX
CC The invention relates to novel human polynucleotides (AAI99513-AAI99538)
CC and the encoded proteins (AAM99915-AAM99934) which are useful for
CC preventing, treating or ameliorating medical conditions e.g. by protein
CC or gene therapy. The genes are isolated from a range of human tissues
CC disclosed in the specification. The nucleic acids, proteins, antibodies
CC and (ant)agonists are useful in the diagnosis, treatment and prevention
CC of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases
CC e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as
CC viral, bacterial, fungal and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1280 BP; 345 A; 277 C; 257 G; 398 T; 3 other;

Query Match 37.4%; Score 25.4; DB 22; Length 1280;
Best Local Similarity 55.9%; Pred. No. 10;
Matches 33; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

Qy 2 CTTCCAACCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGT 60
Db 739 CTTCTGAACAACTTAGGAGAAACTTTTGATCTTCAGCTTGGTAGATTTAATTGCCCAGT 681

RESULT 13
AAL44063
ID AAL44063 standard; cDNA; 2257 BP.
XX
AC AAL44063;
XX
DT 27-SEP-2002 (first entry)
XX
DE Human genset metabolic gene (GMG-7A) cDNA sequence.
XX
KW Human; gene; ss; gene therapy; genset metabolic gene; GMG-7A; GMG-7B;
KW GMG-8; GMG-9; GMG-10; GMG-11; metabolic-related disorder; obesity;
KW impaired glucose tolerance; insulin resistance; Syndrome X;
KW Type II diabetes; hyperlipidaemia; atherosclerosis; hypertension;
KW heart disease; cardiac insufficiency; coronary insufficiency;
KW high blood pressure; insulin sensitiser;
KW non-insulin dependent diabetes mellitus.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 37..2169
FT /*tag= a
FT /product= "Human GMG-7A protein"
XX
PN WO200255694-A2.
XX
PD 18-JUL-2002.
XX
PF 15-JAN-2002; 2002WO-IB01215.
XX
PR 16-JAN-2001; 2001US-262235P.
XX
PA (GEST) GENSET.
XX
PI Erickson MR, Bour BA, Bihain B, Tanaka H;
XX
DR WPI; 2002-557821/59.
DR P-PSDB; AAO15420.
XX
PT Treating or preventing a metabolic-related disease or disorder, e.g.
PT obesity, impaired glucose tolerance, insulin resistance, Syndrome X, or
PT Type II diabetes, comprises administering Genset Metabolic Genes -
XX
PS Disclosure; Page 113-116; 128pp; English.
XX
CC The invention comprises the amino acid and coding sequences of six human
CC genset metabolic genes (GMG-7A, GMG-7B, GMG-8, GMG-9, GMG-10 and GMG-11).
CC The GMG DNA and protein sequences of the invention are useful for
CC treating or preventing metabolic-related disorders, such as: obesity;
CC impaired glucose tolerance; insulin resistance; Syndrome X; Type II
CC diabetes; hyperlipidaemia; atherosclerosis; hypertension; and heart

CC diseases (e.g. cardiac insufficiency, coronary insufficiency or high
CC blood pressure). The GMG DNA and protein sequences of the invention may
CC also be used as insulin sensitisers - for improving insulin sensitivity
CC in persons with non-insulin dependent diabetes mellitus. The present cDNA
CC sequence encodes the human GMG-7A protein.
XX
SQ Sequence 2257 BP; 728 A; 522 C; 482 G; 525 T; 0 other;

Query Match 37.4%; Score 25.4; DB 24; Length 2257;
Best Local Similarity 55.9%; Pred. No. 12;
Matches 33; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

Qy 2 CTTCCAACCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGT 60
Db 1849 CTTCTGAACAACTTAGGAGAAACTTTTGATCTTCAGCTTGGTAGATTTAATTGCCCAGT 1907

RESULT 14
AAK52451
ID AAK52451 standard; cDNA; 2454 BP.
XX
AC AAK52451;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1980.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM79318.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 1; Page 4386-4387; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 2454 BP; 782 A; 574 C; 532 G; 563 T; 3 other;

Query Match 37.4%; Score 25.4; DB 22; Length 2454;
Best Local Similarity 55.9%; Pred. No. 12;
Matches 33; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

Qy 2 CTTCCAACCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGT 60
Db 2175 CTTCTGAACAACTTAGGAGAAACTTTTGATCTTCAGCTTGGTAGATTTAATTGCCCAGT 2233

RESULT 15
AAK52452
ID AAK52452 standard; cDNA; 2454 BP.
XX
AC AAK52452;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1981.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM79319.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 1; Page 4387-4388; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 2454 BP; 782 A; 574 C; 532 G; 563 T; 3 other;

Query Match 37.4%; Score 25.4; DB 22; Length 2454;
Best Local Similarity 55.9%; Pred. No. 12;
Matches 33; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

```
Qy       2 CTTCCAACCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGT 60
           ||||  | | || ||| ||||| |||| |  ::| ||:  |||| ::  |  | |  ||
Db    2175 CTTCTGAACAACTTAGGAGAAACTTTTGATCTTCAGCTTGGTAGATTTAATTGCCCAGT 2233
```

Search completed: March 17, 2003, 10:50:28
Job time : 393.887 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 10:23:12 ; Search time 2403.76 Seconds
(without alignments)
458.154 Million cell updates/sec

Title: US-09-836-439-1
Perfect score: 68
Sequence: 1 ccttccaacctacgtagcag.....ggaagggcgcgttttcgcgc 68

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	27	39.7	818	12	BG188346	BG188346 RST7364 A
	2	27	39.7	1101	17	CNS0034I	AL063844 Drosophil
c	3	25.8	37.9	741	13	BI739136	BI739136 603361404
c	4	25.8	37.9	747	13	BI733738	BI733738 603352352
c	5	25.8	37.9	899	14	BQ946406	BQ946406 AGENCOURT
c	6	25.8	37.9	923	14	BQ937919	BQ937919 AGENCOURT
c	7	25.8	37.9	943	12	BG262012	BG262012 602373843
c	8	25.8	37.9	961	12	BG342737	BG342737 602374557
c	9	25.8	37.9	1101	12	BG295440	BG295440 602392754
c	10	25.6	37.6	288	9	AU163129	AU163129 AU163129
c	11	25.6	37.6	297	9	AU096979	AU096979 AU096979
c	12	25.6	37.6	300	10	BE401698	BE401698 CNW02EL01
	13	25.6	37.6	349	9	AU070730	AU070730 AU070730
c	14	25.6	37.6	352	9	AU057142	AU057142 AU057142
c	15	25.6	37.6	356	9	AU162789	AU162789 AU162789
	16	25.6	37.6	358	9	AU070946	AU070946 AU070946
c	17	25.6	37.6	398	10	BE229338	BE229338 98MJ0005
	18	25.6	37.6	463	9	AU057241	AU057241 AU057241
c	19	25.6	37.6	465	9	AU164976	AU164976 AU164976
c	20	25.6	37.6	507	10	BE230696	BE230696 99AS93 Ri
	21	25.6	37.6	573	17	AQ803248	AQ803248 HS_3154_A
	22	25.6	37.6	606	17	AQ049317	AQ049317 cLM-13f4-
c	23	25.6	37.6	909	9	AL566335	AL566335 AL566335
	24	25.4	37.4	432	12	BF149104	BF149104 35_11 Hum
c	25	25.4	37.4	464	10	AW377397	AW377397 QV0-CT022
	26	25.4	37.4	466	9	AA625142	AA625142 af70c02.r
	27	25.4	37.4	472	12	BF048056	BF048056 dc88h10.y
	28	25.4	37.4	494	14	BM706099	BM706099 UI-E-DW0-
	29	25.4	37.4	510	10	AW022120	AW022120 df33b07.y
c	30	25.4	37.4	574	14	BM969655	BM969655 UI-CF-DU1
	31	25.4	37.4	576	14	BM696870	BM696870 UI-E-DW0-
	32	25.4	37.4	581	14	BM693772	BM693772 UI-E-DW1-
	33	25.4	37.4	582	14	BM696816	BM696816 UI-E-DW0-
	34	25.4	37.4	582	17	AZ250774	AZ250774 RPCI-23-6
	35	25.4	37.4	625	12	BG120479	BG120479 602347192
	36	25.4	37.4	632	17	AG001612	AG001612 Homo sapi
	37	25.4	37.4	674	17	AG105983	AG105983 Pan trogl
	38	25.4	37.4	675	13	BI596522	BI596522 603243571
c	39	25.4	37.4	719	14	BQ009544	BQ009544 UI-H-ED1-
	40	25.4	37.4	739	12	BG705003	BG705003 602688996
	41	25.4	37.4	777	12	BG213748	BG213748 RST33367
	42	25.4	37.4	779	12	BG201727	BG201727 RST21069
	43	25.4	37.4	784	12	BG198633	BG198633 RST17898
	44	25.4	37.4	812	12	BG199159	BG199159 RST18438
	45	25.4	37.4	824	12	BG219546	BG219546 RST39305

ALIGNMENTS

RESULT 1
BG188346
LOCUS BG188346 818 bp mRNA linear EST 21-APR-2001
DEFINITION RST7364 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG188346
VERSION BG188346.1 GI:13710033
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 818)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J. and Ducar,M.
TITLE Creation of genome-wide protein expression libraries using random activation of gene expression
JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE 21227151
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 440.
FEATURES Location/Qualifiers

```
     source          1. .818
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Athersys RAGE Library"
                     /cell_line="HT1080"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
BASE COUNT      218 a    197 c    181 g    222 t
ORIGIN

  Query Match             39.7%;  Score 27;  DB 12;  Length 818;
  Best Local Similarity   57.6%;  Pred. No. 17;
  Matches   34;  Conservative    5;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       2 CTTCCAACCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGT 60
           ||||  | | || ||| ||||| |||| |  ::| ||:  |||| ::  |  |||  ||
Db     464 CTTCTGAACAACTTAGGAGAAACTTTTGATCTTCAGCTTGGTAGATTTAATTGGCCAGT 522


RESULT 2
CNS0034I
LOCUS       CNS0034I                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR07N08 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063844
VERSION     AL063844.1  GI:4941600
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR07N08"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      349 a    179 c    216 g    325 t     32 others
ORIGIN

  Query Match             39.7%;  Score 27;  DB 17;  Length 1101;
  Best Local Similarity   58.8%;  Pred. No. 19;
  Matches   30;  Conservative    6;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      15 TAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGGCGCGTTTTCG 65
           | ||| |||||||||||: : : |:|   |  ::||||| |||  | || |
Db     693 TTGCATAAAGTTTTTACTGTTTACTAGTCATTTTGGAAGTGCGTTTGTTGG 743


RESULT 3
BI739136/c
LOCUS       BI739136                741 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603361404F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5368368 5',
            mRNA sequence.
ACCESSION   BI739136
VERSION     BI739136.1  GI:15716149
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 741)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11937  row: h  column: 01
            High quality sequence stop: 729.
FEATURES             Location/Qualifiers
     source          1. .741
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5368368"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      210 a    177 c    194 g    160 t
ORIGIN

  Query Match             37.9%;  Score 25.8;  DB 13;  Length 741;
  Best Local Similarity   69.0%;  Pred. No. 47;
  Matches   20;  Conservative    7;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      30 ACUUUCUGCUACGTAGGUUGGAAGGGCGC 58
           ||:::|:||:|||||||::|||||| | |
Db     368 ACTTTCTGCTACGTAGGTTGGAAGGCCAC 340


RESULT 4
BI733738/c
LOCUS       BI733738                747 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603352352F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5359908 5',
            mRNA sequence.
ACCESSION   BI733738
VERSION     BI733738.1  GI:15710751
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11915  row: g  column: 13
            High quality sequence stop: 734.
FEATURES              Location/Qualifiers
     source           1. .747
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:5359908"
                      /clone_lib="NIH_MGC_94"
                      /tissue_type="retina"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                      Average insert size 3.3 kb. Library enriched for
                      full-length clones and constructed by Life Technologies.
                      Note: this is a NIH_MGC Library."
BASE COUNT      218 a    169 c    198 g    162 t
ORIGIN

  Query Match             37.9%;  Score 25.8;  DB 13;  Length 747;
  Best Local Similarity   69.0%;  Pred. No. 47;
  Matches   20;  Conservative    7;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      30 ACUUUCUGCUACGTAGGUUGGAAGGGCGC 58
           ||:::|:||:|||||||::|||||| | |
Db     183 ACTTTCTGCTACGTAGGTTGGAAGGCCAC 155


RESULT 5
BQ946406/c
LOCUS       BQ946406                899 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8924053 NIH_MGC_94 Mus musculus cDNA clone IMAGE:6467512
            5', mRNA sequence.
ACCESSION   BQ946406
VERSION     BQ946406.1  GI:22361884
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 899)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13993  row: i  column: 17
            High quality sequence stop: 614.
FEATURES              Location/Qualifiers
     source           1. .899
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:6467512"
                      /clone_lib="NIH_MGC_94"
                      /tissue_type="retina"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                      Average insert size 3.3 kb. Library enriched for
                      full-length clones and constructed by Life Technologies.
                      Note: this is a NIH_MGC Library."
BASE COUNT      248 a    208 c    231 g    207 t      5 others
ORIGIN

  Query Match             37.9%;  Score 25.8;  DB 14;  Length 899;
  Best Local Similarity   69.0%;  Pred. No. 50;
  Matches   20;  Conservative    7;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      30 ACUUUCUGCUACGTAGGUUGGAAGGGCGC 58
           ||:::|:||:|||||||::|||||| | |
Db     532 ACTTTCTGCTACGTAGGTTGGAAGGCCAC 504


RESULT 6
BQ937919/c
LOCUS       BQ937919                923 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8932027 NIH_MGC_94 Mus musculus cDNA clone IMAGE:6465857
            5', mRNA sequence.
ACCESSION   BQ937919
VERSION     BQ937919.1  GI:22353397
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 923)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13989  row: d  column: 18
            High quality sequence stop: 666.
FEATURES              Location/Qualifiers
     source           1. .923
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:6465857"
                      /clone_lib="NIH_MGC_94"
                      /tissue_type="retina"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                      Average insert size 3.3 kb. Library enriched for
                      full-length clones and constructed by Life Technologies.
                      Note: this is a NIH_MGC Library."
BASE COUNT      251 a    224 c    233 g    214 t      1 others
ORIGIN

  Query Match             37.9%;  Score 25.8;  DB 14;  Length 923;
  Best Local Similarity   69.0%;  Pred. No. 50;
  Matches   20;  Conservative    7;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      30 ACUUUCUGCUACGTAGGUUGGAAGGGCGC 58
           ||:::|:||:|||||||::|||||| | |
Db     667 ACTTTCTGCTACGTAGGTTGGAAGGCCAC 639


RESULT 7
BG262012/c
LOCUS       BG262012                943 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  602373843F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4481291 5',
            mRNA sequence.
ACCESSION   BG262012
VERSION     BG262012.1  GI:12771828
```

```
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 943)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10316  row: f  column: 12
            High quality sequence stop: 694.
FEATURES             Location/Qualifiers
     source          1. .943
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4481291"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      269 a    230 c    256 g    188 t
ORIGIN

  Query Match             37.9%;  Score 25.8;  DB 12;  Length 943;
  Best Local Similarity   69.0%;  Pred. No. 51;
  Matches   20;  Conservative    7;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      30 ACUUUCUGCUACGTAGGUUGGAAGGGCGC 58
           ||:::|:||:|||||||::|||||| | |
Db     471 ACTTTCTGCTACGTAGGTTGGAAGGCCAC 443


RESULT 8
BG342737/c
LOCUS       BG342737                961 bp    mRNA    linear   EST 27-FEB-2001
DEFINITION  602374557F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4481943 5',
            mRNA sequence.
ACCESSION   BG342737
VERSION     BG342737.1  GI:13149175
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 961)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10318  row: a  column: 16
            High quality sequence stop: 653.
FEATURES             Location/Qualifiers
     source          1. .961
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4481943"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      288 a    218 c    254 g    201 t
ORIGIN

  Query Match             37.9%;  Score 25.8;  DB 12;  Length 961;
  Best Local Similarity   69.0%;  Pred. No. 51;
  Matches   20;  Conservative    7;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      30 ACUUUCUGCUACGTAGGUUGGAAGGGCGC 58
           ||:::|:||:|||||||::|||||| | |
Db      31 ACTTTCTGCTACGTAGGTTGGAAGGCCAC 3


RESULT 9
BG295440/c
LOCUS       BG295440                1101 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602392754F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4504476 5',
            mRNA sequence.
ACCESSION   BG295440
VERSION     BG295440.1  GI:13057077
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10376  row: l  column: 13
            High quality sequence stop: 735.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4504476"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      329 a    279 c    289 g    204 t
ORIGIN

  Query Match             37.9%;  Score 25.8;  DB 12;  Length 1101;
  Best Local Similarity   69.0%;  Pred. No. 54;
  Matches   20;  Conservative    7;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      30 ACUUUCUGCUACGTAGGUUGGAAGGGCGC 58
           ||:::|:||:|||||||::|||||| | |
```

```
Db      76 ACTTTCTGCTACGTAGGTTGGAAGGCCAC 48


RESULT 10
AU163129/c
LOCUS       AU163129                288 bp    mRNA    linear   EST 03-APR-2002
DEFINITION  AU163129 Rice green shoot Oryza sativa (japonica cultivar-group)
            cDNA clone S16022, mRNA sequence.
ACCESSION   AU163129
VERSION     AU163129.1  GI:11026528
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group).
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 288)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from green shoot (2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .288
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="S16022"
                     /clone_lib="Rice green shoot"
                     /note="Green shoot (8 days old)"
BASE COUNT       83 a     79 c     68 g     57 t      1 others
ORIGIN

  Query Match             37.6%;  Score 25.6;  DB 9;  Length 288;
  Best Local Similarity   60.4%;  Pred. No. 41;
  Matches   29;  Conservative    5;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       8 ACCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGG 55
           || ||||||| ||  ||| |||| ::  :| :||||| |:  | | ||
Db     210 ACGTACGTAGTAGCCAGTATTTATTTAGTGGTACGTACGTACGCATGG 163


RESULT 11
AU096979/c
LOCUS       AU096979                297 bp    mRNA    linear   EST 03-APR-2002
DEFINITION  AU096979 Rice green shoot Oryza sativa (japonica cultivar-group)
            cDNA clone S16784, mRNA sequence.
ACCESSION   AU096979
VERSION     AU096979.1  GI:8859661
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group).
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 297)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from green shoot (2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .297
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="S16784"
                     /clone_lib="Rice green shoot"
                     /note="Green shoot (8 days old)"
BASE COUNT       83 a     84 c     72 g     56 t      2 others
ORIGIN

  Query Match             37.6%;  Score 25.6;  DB 9;  Length 297;
  Best Local Similarity   60.4%;  Pred. No. 41;
  Matches   29;  Conservative    5;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       8 ACCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGG 55
           || ||||||| ||  ||| |||| ::  :| :||||| |:  | | ||
Db     219 ACGTACGTAGTAGCCAGTATTTATTTAGTGGTACGTACGTACGCATGG 172


RESULT 12
BE401698/c
LOCUS       BE401698                300 bp    mRNA    linear   EST 21-JUL-2000
DEFINITION  CNW02EL010 ITEC CNW Wheat Powdery Mildew Resistant Library Triticum
            aestivum cDNA clone CNW02EL010, mRNA sequence.
ACCESSION   BE401698
VERSION     BE401698.1  GI:9361166
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Jia J
            Key Lab. of Crop Germplasm & Biotechnology
            Inst. of Crop Germplasm Resources
            Beijing 100081 PR CHINA
            Tel: 86 1 62186623
            Fax: 86 1 62186629
            Email: jizeng@mimi.cnc.ac.cn
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Triticum aestivum"
                     /cultivar="Powdery Mildew Resistant line"
                     /db_xref="taxon:4565"
                     /clone="CNW02EL010"
                     /clone_lib="ITEC CNW Wheat Powdery Mildew Resistant
                     Library"
                     /tissue_type="leaf"
                     /dev_stage="seedling, challenged with powdery mildew
                     strain"
                     /note="Vector: Lambda TriplEX; Site_1: SfiI; Primer used:
                     5'-TCCGAGATCTGGACGAGC-3'  500 bp average insert size."
BASE COUNT       84 a     81 c     87 g     48 t
ORIGIN

  Query Match             37.6%;  Score 25.6;  DB 10;  Length 300;
  Best Local Similarity   60.4%;  Pred. No. 41;
  Matches   29;  Conservative    5;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       8 ACCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGG 55
```

```
          || ||||||| ||  ||| |||| ::  :| :|||||| |:  | | ||
Db    260 ACGTACGTAGTAGCCAGTATTTATTTAGTGGTACGTACGTACGCATGG 213


RESULT 13
AU070730
LOCUS       AU070730                349 bp    mRNA    linear   EST 02-APR-2002
DEFINITION  AU070730 Rice cDNA from young root Oryza sativa (japonica
            cultivar-group) cDNA clone R10161_2A, mRNA sequence.
ACCESSION   AU070730
VERSION     AU070730.1  GI:5038620
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group).
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from young root
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .349
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="R10161_2A"
                     /clone_lib="Rice cDNA from young root"
                     /tissue_type="young root"
BASE COUNT       58 a     98 c     96 g     97 t
ORIGIN

  Query Match             37.6%;  Score 25.6;  DB 9;  Length 349;
  Best Local Similarity   60.4%;  Pred. No. 43;
  Matches   29;  Conservative    5;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       8 ACCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGG 55
           || ||||||| ||  ||| |||| ::  :| :|||||| |:  | | ||
Db      85 ACGTACGTAGTAGCCAGTATTTATTTAGTGGTACGTACGTACGCATGG 132


RESULT 14
AU057142/c
LOCUS       AU057142                352 bp    mRNA    linear   EST 01-APR-2002
DEFINITION  AU057142 Oryza sativa mature leaf Nipponbare Oryza sativa (japonica
            cultivar-group) cDNA clone S21188_1A, mRNA sequence.
ACCESSION   AU057142
VERSION     AU057142.1  GI:4716026
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group).
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from mature leaf
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .352
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="S21188_1A"
                     /clone_lib="Oryza sativa mature leaf Nipponbare"
                     /tissue_type="mature leaf"
BASE COUNT       97 a     97 c    100 g     58 t
ORIGIN

  Query Match             37.6%;  Score 25.6;  DB 9;  Length 352;
  Best Local Similarity   60.4%;  Pred. No. 44;
  Matches   29;  Conservative    5;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       8 ACCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGG 55
           || ||||||| ||  ||| |||| ::  :| :|||||| |:  | | ||
Db     268 ACGTACGTAGTAGCCAGTATTTATTTAGTGGTACGTACGTACGCATGG 221


RESULT 15
AU162789/c
LOCUS       AU162789                356 bp    mRNA    linear   EST 03-APR-2002
DEFINITION  AU162789 Rice mature leaf Oryza sativa (japonica cultivar-group)
            cDNA clone S21904, mRNA sequence.
ACCESSION   AU162789
VERSION     AU162789.1  GI:11026188
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group).
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 356)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from mature leaf (2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .356
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="S21904"
                     /clone_lib="Rice mature leaf"
                     /tissue_type="mature leaf"
BASE COUNT       98 a     99 c     99 g     58 t      2 others
ORIGIN

  Query Match             37.6%;  Score 25.6;  DB 9;  Length 356;
  Best Local Similarity   60.4%;  Pred. No. 44;
  Matches   29;  Conservative    5;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       8 ACCTACGTAGCAGAAAGTTTTTACUUUCUGCUACGTAGGUUGGAAGGG 55
           || ||||||| ||  ||| |||| ::  :| :|||||| |:  | | ||
Db     272 ACGTACGTAGTAGCCAGTATTTATTTAGTGGTACGTACGTACGCATGG 225


Search completed: March 17, 2003, 13:09:07
Job time : 2409.76 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:43:42 ; Search time 230.108 Seconds
(without alignments)
3161.870 Million cell updates/sec

Title: US-09-836-439-2
Perfect score: 25
Sequence: 1 ccatgtgaccattaggaaatgagag 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		23.4	93.6	85	10	MMHIF1AS04	AF004144 Mus muscu
2		23.4	93.6	538	9	HSHIF1A03	AF050117 Homo sapi
3		23.4	93.6	2481	6	AX451938	AX451938 Sequence
4		23.4	93.6	2481	6	AX481424	AX481424 Sequence
5		23.4	93.6	2509	9	AF304431	AF304431 Homo sapi
6		23.4	93.6	2522	9	AF208487	AF208487 Homo sapi
7		23.4	93.6	2528	6	AX230580	AX230580 Sequence
8		23.4	93.6	2530	9	AF207601	AF207601 Homo sapi
9		23.4	93.6	2537	9	AF207602	AF207602 Homo sapi
10		23.4	93.6	2711	10	RNHYPFAC1	Y09507 R.norvegicu
11		23.4	93.6	3551	9	AB073325	AB073325 Homo sapi
12		23.4	93.6	3678	9	HSU22431	U22431 Human hypox
13		23.4	93.6	3718	10	AF057308	AF057308 Rattus no
14		23.4	93.6	3746	6	AX306008	AX306008 Sequence
15		23.4	93.6	3746	10	MMHIF1ALP	X95580 M.musculus
16		23.4	93.6	3867	10	MMU59496	U59496 Mus musculu
17		23.4	93.6	3927	6	AX321214	AX321214 Sequence
18		23.4	93.6	3933	9	HSU29165	U29165 Human MOP1
19		23.4	93.6	3942	9	BC012527	BC012527 Homo sapi
20		23.4	93.6	3973	10	AF003695	AF003695 Mus muscu
21		23.4	93.6	4183	10	BC026139	BC026139 Mus muscu
22		23.4	93.6	10355	6	AX230582	AX230582 Sequence
23		23.4	93.6	20463	10	MMY09085	Y09085 Mus musculu
24	c	23.4	93.6	169792	2	AC104313	AC104313 Rattus no
25		23.4	93.6	188107	9	CNS01DWE	AL137129 Human chr
26	c	23.4	93.6	228786	2	AC124712	AC124712 Mus muscu
27		21.8	87.2	2551	4	AB018398	AB018398 Bos tauru
28		20.4	81.6	171838	2	AC105321	AC105321 Mus muscu
29		20.2	80.8	3006	5	AF402782	AF402782 Fundulus
30		20.2	80.8	3076	5	AF212989	AF212989 Coturnix
31	c	20.2	80.8	117771	9	AC112906	AC112906 Homo sapi
32		20.2	80.8	181825	2	AC068191	AC068191 Homo sapi
33		19.8	79.2	141762	9	HS198C21	AL034400 Human DNA
34	c	19.8	79.2	148337	2	AC021509	AC021509 Homo sapi
35	c	19.8	79.2	165840	9	AC068112	AC068112 Homo sapi
36		19.8	79.2	187277	9	AC022968	AC022968 Homo sapi
37	c	19.8	79.2	200000	9	AP000494	AP000494 Homo sapi
38	c	19.8	79.2	200567	2	AP001587	AP001587 Homo sapi
39	c	19.8	79.2	203540	2	AP002396	AP002396 Homo sapi
40		19.4	77.6	1596	8	YSCL30RP	M34387 S.cerevisia
41	c	19.4	77.6	2110	8	SCYGR148C	Z72933 S.cerevisia
42	c	19.4	77.6	2679	8	SCYGR149W	Z72934 S.cerevisia
43	c	19.4	77.6	26677	8	SCCVIIRA	X85807 S.cerevisia
44		19.4	77.6	201802	2	AC124473	AC124473 Mus muscu
45		19.4	77.6	264009	2	AC115294	AC115294 Mus muscu

ALIGNMENTS

RESULT 1
MMHIF1AS04
LOCUS MMHIF1AS04 85 bp DNA linear ROD 25-NOV-1997
DEFINITION Mus musculus hypoxia-inducible factor 1 alpha (Hif1a) gene, exon 4.
ACCESSION AF004144
VERSION AF004144.1 GI:2197137
KEYWORDS .
SEGMENT 4 of 15
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 85)
AUTHORS Luo,G., Gu,Y.Z., Jain,S., Chan,W.K., Carr,K.M., Hogenesch,J.B. and
Bradfield,C.A.

```
  TITLE     Molecular characterization of the murine Hif-1 alpha locus
  JOURNAL   Gene Expr. 6 (5), 287-299 (1997)
  MEDLINE   98034461
   PUBMED   9368100
REFERENCE   2  (bases 1 to 85)
  AUTHORS   Luo,G., Gu,Y.-Z., Jain,S., Chan,W.K., Carr,K.M., Hogenesch,J.B. and
            Bradfield,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1997) Oncology, University of Wisconsin-Madison,
            1400 University Ave., Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .85
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     exon            1. .85
                     /gene="Hif1a"
                     /number=4
BASE COUNT       29 a     14 c     18 g     24 t
ORIGIN

  Query Match           93.6%;  Score 23.4;  DB 10;  Length 85;
  Best Local Similarity   96.0%;  Pred. No. 0.43;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||||| |||||||||||||
Db      40 CCATGTGACCATGAGGAAATGAGAG 64


RESULT 2
HSHIF1A03
LOCUS       HSHIF1A03               538 bp    DNA     linear   PRI 26-OCT-1998
DEFINITION  Homo sapiens hypoxia-inducible factor 1 alpha subunit (HIF1A) gene,
            exons 3 and 4.
ACCESSION   AF050117
VERSION     AF050117.1  GI:3790523
KEYWORDS    .
SEGMENT     3 of 13
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 538)
  AUTHORS   Iyer,N.V., Leung,S.W. and Semenza,G.L.
  TITLE     The human hypoxia-inducible factor 1alpha gene: HIF1A structure and
            evolutionary conservation
  JOURNAL   Genomics 52 (2), 159-165 (1998)
  MEDLINE   99000835
   PUBMED   9782081
REFERENCE   2  (bases 1 to 538)
  AUTHORS   Iyer,N.V., Leung,S.W. and Semenza,G.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1998) Departments of Pediatrics and Medicine,
            Institute of Genetic Medicine, Johns Hopkins University School of
            Medicine, 600 N. Wolfe St, Baltimore, MD 21287-3914, USA
FEATURES             Location/Qualifiers
     source          1. .538
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q21-q24"
     intron          <1. .67
                     /gene="HIF1A"
                     /number=2
     exon            68. .213
                     /gene="HIF1A"
                     /number=3
     intron          214. .297
                     /gene="HIF1A"
                     /number=3
     exon            298. .382
                     /gene="HIF1A"
                     /number=4
     intron          383. .>538
                     /gene="HIF1A"
                     /number=4
BASE COUNT      170 a     69 c     89 g    210 t
ORIGIN

  Query Match           93.6%;  Score 23.4;  DB 9;  Length 538;
  Best Local Similarity   96.0%;  Pred. No. 0.41;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||||| |||||||||||||
Db     337 CCATGTGACCATGAGGAAATGAGAG 361


RESULT 3
AX451938
LOCUS       AX451938                2481 bp    DNA     linear   PAT 03-JUL-2002
DEFINITION  Sequence 3 from Patent WO0212326.
ACCESSION   AX451938
VERSION     AX451938.1  GI:21698761
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Poellinger,L., Pereira,T. and Ruas,J.
  TITLE     Mechanism of conditional regulation of the hypoxia-inducible
            factor-1 by the von hippel-lindau tumor suppressor protein
  JOURNAL   Patent: WO 0212326-A 3 14-FEB-2002;
            Angiogenetics Sweden AB (SE)
FEATURES             Location/Qualifiers
     source          1. .2481
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             1. .2481
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD37798.1"
                     /db_xref="GI:21698762"
                     /translation="MEGAGGANDKKKISSERRKEKSRDAARSRRSKESEVFYELAHQL
                     PLPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLDIEDDMKAQMNCFYLKALDGFVMV
                     LTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEMREMLTHRNGLVKKGKE
                     QNTQRSFFLRMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYDTNSNQPQCGYKKPPMT
                     CLVLICEPIPHPSNIEIPLDSKTFLSRHSLDMKFSYCDERITELMGYEPEELLGRSIY
                     EYYHALDSDHLTKTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQATVIYNTKNSQPQC
                     IVCVNYVVSGIIQHDLIFSLQQTECVLKPVESSDMKMTQLFTKVESEDTSSLFDKLKK
                     EPDALTLLAPAAGDTIISLDFGSNDTETDDQQLEEVPLYNDVMLPSPNEKLQNTNLAM
                     SPLPTAETPKPLRSSADPALNQEVALKLEPNPESLELSFTMPQIQDQTPSPSDGSTRQ
                     SSPEPNSPSEYCFYVDSDMVNEFKLELVEKLFAEDTEAKNPFSTQDTDLDLEMLAPYI
                     PMDDDFQLRSFDQLSPLESSSASPESASPQSTVTVFQQTQIQEPTANATTTTATTDEL
                     KTVTKDRMEDIKILIASPSPTHIHKETTSATSSPYRDTQSRTASPNRAGKGVIEQTEK
                     SHPRSPNVLSVALSQRTTVPEEELNPKILALQNAQRKRKMEHDGSLFQAVGIGTLLQQ
                     PDDHAATTSLSWKRVKGCKSSEQNGMEQKTIILIPSDLACRLLGQSMDESGLPQLTSY
                     DCEVNAPIQGSRNLLQGEELLRALDQVN"
BASE COUNT      829 a    512 c    500 g    640 t
ORIGIN

  Query Match           93.6%;  Score 23.4;  DB 6;  Length 2481;
  Best Local Similarity   96.0%;  Pred. No. 0.38;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||||| |||||||||||||
Db     412 CCATGTGACCATGAGGAAATGAGAG 436


RESULT 4
AX481424
LOCUS       AX481424                2481 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 38 from Patent WO02055693.
ACCESSION   AX481424
```

```
VERSION     AX481424.1  GI:22316338
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kreutzer,R., Limmer,S., Rost,S. and Hadwiger,P.
  TITLE     Method for inhibiting the expression of a target gene
  JOURNAL   Patent: WO 02055693-A 38 18-JUL-2002;
            Ribopharma AG (DE)
FEATURES             Location/Qualifiers
     source          1. .2481
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      829 a    512 c    500 g    640 t
ORIGIN
```

Query Match 93.6%; Score 23.4; DB 6; Length 2481;
Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 CCATGTGACCATTAGGAAATGAGAG 25
          ||||||||||||| |||||||||||||
Db    412 CCATGTGACCATGAGGAAATGAGAG 436
```

```
RESULT 5
AF304431
LOCUS       AF304431                2509 bp    mRNA    linear   PRI 29-DEC-2000
DEFINITION  Homo sapiens hypoxia-inducible factor 1 alpha subunit (HIF1A) mRNA,
            complete cds.
ACCESSION   AF304431
VERSION     AF304431.1  GI:11995454
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2509)
  AUTHORS   Sun,B., Zhao,H.R., Yu,R.T. and Ni,M.S.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-2000) Department of Neurosurgery, Affiliated
            Hospital of Xuzhou Medical College, Huaihai West Road, Xuzhou,
            Jiangsu 221002, China
FEATURES             Location/Qualifiers
     source          1. .2509
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q21-q24"
                     /tissue_type="glioma"
     gene            1. .2509
                     /gene="HIF1A"
     CDS             1. .2481
                     /gene="HIF1A"
                     /codon_start=1
                     /product="hypoxia-inducible factor 1 alpha subunit"
                     /protein_id="AAG43026.1"
                     /db_xref="GI:11995455"
                     /translation="MEGAGGANDKKKISSERRKEKSRDAARSRRSKESEVFYELAHQL
                     PLPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLDIEDDMKAQMNCFYLKALDGFVMV
                     LTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEMREMLTHRNGLVKKGKE
                     QNTQRSFFLRMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYDTNSNQPQCGYKKPPMT
                     CLVLICEPIPHPSNIEIPLDSKTFLSRHSLDMKFSYCDERITELMGYEPEELLGRSIY
                     EYYHALDSDHLTKTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQATVIYNTKNSQPQC
                     IVCVNYVVSGIIQHDLIFSLQQTECVLKPVESSDMKMTQLFTKVESEDTSSLFDKLKK
                     EPDALTLLAPAAGDTIISLDFGSNDTETDDQQLEEVPLYNDVMLPSPNEKLQNINLAM
                     SPLPTAETPKPLRSSADPALNQEVALKLEPNPESLELSFTMPQIQDQTPSPSDGSTRQ
                     SSPEPNSPSEYCFYVDSDMVNEFKLELVEKLFAEDTEAKNPFSTQDTDLDLEMLAPYI
                     PMDDDFQLRSFDQLSPLESSSASPESASPQSTVTVFQQTQIQEPTANATTTTATTDEL
                     KTVTKDRMEDIKILIASPSPTHIHKETTSATSSPYRDTQSRTASPNRAGKGVIEQTEK
                     SHPRSPNVLSVALSQRTTVPEEELNPKILALQNAQRKRKMEHDGSLFQAVGIGTLLQQ
                     PDDHAATTSLSWKRVKGCKSSEQNGMEQKTIILIPSDLACRLLGQSMDESGLPQLTSY
                     DCEVNAPIQGSRNLLQGEELLRALDQVN"
BASE COUNT      832 a    517 c    502 g    658 t
ORIGIN
```

Query Match 93.6%; Score 23.4; DB 9; Length 2509;
Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 CCATGTGACCATTAGGAAATGAGAG 25
          ||||||||||||| |||||||||||||
Db    412 CCATGTGACCATGAGGAAATGAGAG 436
```

```
RESULT 6
AF208487
LOCUS       AF208487                2522 bp    DNA     linear   PRI 27-DEC-1999
DEFINITION  Homo sapiens hypoxia-inducible factor 1 alpha (HIF1A) gene,
            complete cds.
ACCESSION   AF208487
VERSION     AF208487.1  GI:6636337
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2522)
  AUTHORS   Rupert,J.L. and Hochachka,P.W.
  TITLE     HIF1a sequence in the Quechua, a high altitude population
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2522)
  AUTHORS   Rupert,J.L. and Hochachka,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-1999) Zoology, University of British Columbia,
            6270 University Blvd., Vancouver, BC V6T 1Z4, Canada
FEATURES             Location/Qualifiers
     source          1. .2522
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q21-q24"
                     /cell_line="GM11198 (Coriell)"
                     /cell_type="lymphoblast"
                     /note="derived from a high altitude native (Quechua)"
     gene            11. .2491
                     /gene="HIF1A"
                     /note="HIF1-alpha"
     CDS             11. .2491
                     /gene="HIF1A"
                     /note="transcription factor; hypoxia-inducible factor 1a"
                     /codon_start=1
                     /product="hypoxia-inducible factor 1 alpha"
                     /protein_id="AAF20149.1"
                     /db_xref="GI:6636338"
                     /translation="MEGAGGANDKKKISSERRKEKSRDAARSRRSKESEVFYELAHQL
                     PLPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLDIEDDMKAQMNCFYLKALDGFVMV
                     LTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEMREMLTHRNGLVKKGKE
                     QNTQRSFFLRMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYDTNSNQPQCGYKKPPMT
                     CLVLICEPIPHPSNIEIPLDSKTFLSRHSLDMKFSYCDERITELMGYEPEELLGRSIY
                     EYYHALDSDHLTKTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQATVIYNTKNSQPQC
                     IVCVNYVVSGIIQHDLIFSLQQTECVLKPVESSDMKMTQLFTKVESEDTSSLFDKLKK
                     EPDALTLLAPAAGDTIISLDFGSNDTETDDQQLEEVPLYNDVMLPSPNEKLQNINLAM
                     SPLPTAETPKPLRSSADPALNQEVALKLEPNPESLELSFTMPQIQDQTPSPSDGSTRQ
                     SSPEPNSPSEYCFYVDSDMVNEFKLELVEKLFAEDTEAKNPFSTQDTDLDLEMLAPYI
                     PMDDDFQLRSFDQLSPLESSSASPESASPQSTVTVFQQTQIQEPTANATTTTATTDEL
                     KTVTKDRMEDIKILIASPSPTHIHKETTSATSSPYRDTQSRTASPNRAGKGVIEQTEK
                     SHPRSPNVLSVALSQRTTVPEEELNPKILALQNAQRKRKMEHDGSLFQAVGIGTLLQQ
                     PDDHAATTSLSWKRVKGCKSSEQNGMEQKTIILIPSDLACRLLGQSMDESGLPQLTSY
                     DCEVNAPIQGSRNLLQGEELLRALDQVN"
BASE COUNT      835 a    523 c    504 g    660 t
ORIGIN
```

Query Match 93.6%; Score 23.4; DB 9; Length 2522;

```
  Best Local Similarity   96.0%;  Pred. No. 0.38;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||| ||||||||||||
Db     422 CCATGTGACCATGAGGAAATGAGAG 446


RESULT 7
AX230580
LOCUS       AX230580               2528 bp    DNA     linear   PAT 11-SEP-2001
DEFINITION  Sequence 2 from Patent WO0162965.
ACCESSION   AX230580
VERSION     AX230580.1  GI:15592425
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2528)
  AUTHORS   Kingsman,A.J.
  TITLE     Differential expression screening method
  JOURNAL   Patent: WO 0162965-A 2 30-AUG-2001;
            Oxford Biomedica (UK) Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .2528
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      839 a    528 c    513 g    648 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 6;  Length 2528;
  Best Local Similarity   96.0%;  Pred. No. 0.38;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||| ||||||||||||
Db     439 CCATGTGACCATGAGGAAATGAGAG 463


RESULT 8
AF207601
LOCUS       AF207601               2530 bp    mRNA    linear   PRI 27-DEC-1999
DEFINITION  Homo sapiens cell-line GM11201 hypoxia-inducible factor 1 alpha
            (HIF1A) mRNA, complete cds.
ACCESSION   AF207601
VERSION     AF207601.1  GI:6636316
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2530)
  AUTHORS   Rupert,J.L. and Hochachka,P.W.
  TITLE     HIF1a sequence in the Quechua, a high altitude population
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2530)
  AUTHORS   Rupert,J.L. and Hochachka,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1999) Zoology, University of British Columbia,
            6270 University Blvd., Vancouver, BC V6T 1Z4, Canada
FEATURES             Location/Qualifiers
     source          1. .2530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q21-q24"
                     /cell_line="GM11201 (Coriell)"
                     /cell_type="lymphoblast"
                     /note="derived from a high altitude native (Quechua)"
     gene            1. .2530
                     /gene="HIF1A"
                     /note="HIF1-alpha"
     CDS             11. .2491
                     /gene="HIF1A"
                     /note="transcription factor; hypoxia-inducible factor 1a"
                     /codon_start=1
                     /product="hypoxia-inducible factor 1 alpha"
                     /protein_id="AAF20139.1"
                     /db_xref="GI:6636317"
                     /translation="MEGAGGANDKKKISSERRKEKSRDAARSRRSKESEVFYELAHQL
                     PLPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLDIEDDMKAQMNCFYLKALDGFVMV
                     LTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEMREMLTHRNGLVKKGKE
                     QNTQRSFFLRMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYDTNSNQPQCGYKKPPMT
                     CLVLICEPIPHPSNIEIPLDSKTFLSRHSLDMKFSYCDERITELMGYEPEELLGRSIY
                     EYYHALDSDHLTKTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQATVIYNTKNSQPQC
                     IVCVNYVVSGIIQHDLIFSLQQTECVLKPVESSDMKMTQLFTKVESEDTSSLFDKLKK
                     EPDALTLLAPAAGDTIISLDFGSNDTETDDQQLEEVPLYNDVMLPSPNEKLQNINLAM
                     SPLPTAETPKPLRSSADPALNQEVALKLEPNPESLELSFTMPQIQDQTPSPSDGSTRQ
                     SSPEPNSPSEYCFYVDSDMVNEFKLELVEKLFAEDTEAKNPFSTQDTDLDLEMLAPYI
                     PMDDDFQLRSFDQLSPLESSSASPESASPQSTVTVFQQTQIQEPTANATTTTATTDEL
                     KTVTKDRMEDIKILIASPSPTHIHKETTSATSSPYRDTQSRTASPNRAGKGVIEQTEK
                     SHPRSPNVLSVALSQRTTVPEEELNPKILALQNAQRKRKMEHDGSLFQAVGIGTLLQQ
                     PDDHAATTSLSWKRVKGCKSSEQNGMEQKTIILIPSDLACRLLGQSMDESGLPQLTSY
                     DCEVNAPIQGSRNLLQGEELLRALDQVN"
BASE COUNT      836 a    524 c    508 g    662 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 9;  Length 2530;
  Best Local Similarity   96.0%;  Pred. No. 0.38;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||| ||||||||||||
Db     422 CCATGTGACCATGAGGAAATGAGAG 446


RESULT 9
AF207602
LOCUS       AF207602               2537 bp    mRNA    linear   PRI 27-DEC-1999
DEFINITION  Homo sapiens cell-line GM11197 hypoxia-inducible factor 1 alpha
            (HIF1A) mRNA, complete cds.
ACCESSION   AF207602
VERSION     AF207602.1  GI:6636318
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2537)
  AUTHORS   Rupert,J.L. and Hochachka,P.W.
  TITLE     HIF1a sequence in the Quechua, a high altitude population
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2537)
  AUTHORS   Rupert,J.L. and Hochachka,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1999) Zoology, University of British Columbia,
            6270 University Blvd., Vancouver, BC V6T 1Z4, Canada
FEATURES             Location/Qualifiers
     source          1. .2537
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q21-q24"
                     /cell_line="GM11197 (Coriell)"
                     /cell_type="lymphoblast"
                     /note="derived from a high altitude native (Quechua)"
     gene            1. .2537
                     /gene="HIF1A"
                     /note="HIF1-alpha"
     CDS             8. .2488
                     /gene="HIF1A"
                     /note="transcription factor; hypoxia-inducible factor 1a"
                     /codon_start=1
                     /product="hypoxia-inducible factor 1 alpha"
                     /protein_id="AAF20140.1"
```

```
                     /db_xref="GI:6636319"
                     /translation="MEGAGGANDKKKISSERRKEKSRDAARSRRSKESEVFYELAHQL
                     PLPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLDIEDDMKAQMNCFYLKALDGFVMV
                     LTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEMREMLTHRNGLVKKGKE
                     QNTQRSFFLRMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYDTNSNQPQCGYKKPPMT
                     CLVLICEPIPHPSNIEIPLDSKTFLSRHSLDMKFSYCDERITELMGYEPEELLGRSIY
                     EYYHALDSDHLTKTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQATVIYNTKNSQPQC
                     IVCVNYVVSGIIQHDLIFSLQQTECVLKPVESSDMKMTQLFTKVESEDTSSLFDKLKK
                     EPDALTLLAPAAGDTIISLDFGSNDTETDDQQLEEVPLYNDVMLPSPNEKLQNINLAM
                     SPLPTAETPKPLRSSADPALNQEVALKLEPNPESLELSFTMPQIQDQTPSPSDGSTRQ
                     SSPEPNSPSEYCFYVDSDMVNEFKLELVEKLFAEDTEAKNPFSTQDTDLDLEMLAPYI
                     PMDDDFQLRSFDQLSPLESSSASPESASPQSTVTVFQQTQIQEPTANATTTTATTDEL
                     KTVTKDRMEDIKILIASPSPTHIHKETTSATSSPYRDTQSRTASPNRAGKGVIEQTEK
                     SHPRSPNVLSVALSQRTTVPEEELNPKILALQNAQRKRKMEHDGSLFQAVGIGTLLQQ
                     PDDHAATTSLSWKRVKGCKSSEQNGMEQKTIILIPSDLACRLLGQSMDESGLPQLTSY
                     DCEVNAPIQGSRNLLQGEELLRALDQVN"
BASE COUNT      838 a    526 c    507 g    666 t
ORIGIN

  Query Match           93.6%;  Score 23.4;  DB 9;  Length 2537;
  Best Local Similarity   96.0%;  Pred. No. 0.38;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCATGTGACCATTAGGAAATGAGAG 25
          ||||||||||||| |||||||||||||
Db    419 CCATGTGACCATGAGGAAATGAGAG 443


RESULT 10
RNHYPFAC1
LOCUS       RNHYPFAC1               2711 bp    mRNA    linear   ROD 26-FEB-2001
DEFINITION  R.norvegicus mRNA for hypoxia-inducible factor 1.
ACCESSION   Y09507
VERSION     Y09507.1  GI:2924301
KEYWORDS    hypoxia-inducible factor 1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2711)
  AUTHORS   Kietzmann,T., Cornesse,Y., Brechtel,K., Modaressi,S. and
            Jungermann,K.
  TITLE     Perivenous expression of the mRNA of the three hypoxia-inducible
            factor alpha-subunits, HIF1alpha, HIF2alpha and HIF3alpha, in rat
            liver
  JOURNAL   Biochem. J. 354 (Pt 3), 531-537 (2001)
  MEDLINE   21134367
   PUBMED   11237857
REFERENCE   2  (bases 1 to 2711)
  AUTHORS   Kietzmann,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1996) T. Kietzmann, Inst. Of Biochem. And Mol.
            Cell Biol., Georg-August-Univ. Goettingen, Humboldtallee 23,
            D-37073 Goettingen, FRG
  REMARK    Revised by [3]
REFERENCE   3  (bases 1 to 2711)
  AUTHORS   Kietzmann,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-1998) T. Kietzmann, Inst. Of Biochem. And Mol.
            Cell Biol., Georg-August-Univ. Goettingen, Humboldtallee 23,
            D-37073 Goettingen, FRG
COMMENT     On Mar 4, 1998 this sequence version replaced gi:2632130.
FEATURES             Location/Qualifiers
     source          1. .2711
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /sex="male"
                     /cell_line="primary hepatocytes"
                     /dev_stage="adult"
     CDS             124. .2601
                     /codon_start=1
                     /product="hypoxia-inducible factor 1"
                     /protein_id="CAA70701.1"
                     /db_xref="GI:2632131"
                     /db_xref="SPTREMBL:O35800"
                     /translation="MEGAGGENEKKKMSSERRKEKSRDAARSRRSKESEVFYELAHQL
                     PLPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLDIEDEMKAQMNCFYLKAPDGFVMV
                     LTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEMREMLTHRNGPVRKGKE
                     QNTQRSFFLRMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYDTSSNQPQCGYKKPPMT
                     CLVLICEPIPHPSNIEIPLDSKTFLSRHSLDMKFSYCDERITELMGYEPEELLGRSIY
                     EYYHALDSDHLTKTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQATVIYNTKDSQPQC
                     IVCVNYVVSGIIQHDLIFSLQQTESVLKPVESSDMKMTQLFTKVESEDTSCLFDKLKK
                     EPDALTLLAPAAGDTIISLDFGSDDTETEDQQLEDVPLYNDVMFPSSNEKLNINLAMS
                     PLPASETPKPLRSSADPALNQEVALKLESSPESLGLSFTMPQIQDQPASPSDGSTRQS
                     SPEPNSPSEYCFDVDSDMVNVFKLELVEKLFAEDTEAKNPFSAQDTDLDLEMLAPYIP
                     MDDDFQLRSFDQLSPLESNSPSPPSVSTVTGFQQTQLQKPTITVTAATTATTTTDES
                     KAVTKDNIEDIKILIASPPSTQVPQEMTTAKASAYSGTHSRTASPDRAGKRVIEKTDK
                     AHPRSLNLSVTLNQRNTVPEEELNPRTIALQNAQRKRKMEHDGSLFQAAGIGTLLQQP
                     GDRAPTMSLSWKRVKGYISSEQDGMEQKTIFLIPSDLACRLLGQSMDESGLPQLTSYD
                     CEVNAPIQGSRNLLQGEELLRALDQVN"
BASE COUNT      798 a    613 c    623 g    677 t
ORIGIN

  Query Match           93.6%;  Score 23.4;  DB 10;  Length 2711;
  Best Local Similarity   96.0%;  Pred. No. 0.38;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCATGTGACCATTAGGAAATGAGAG 25
          ||||||||||||| |||||||||||||
Db    535 CCATGTGACCATGAGGAAATGAGAG 559


RESULT 11
AB073325
LOCUS       AB073325                3551 bp    mRNA    linear   PRI 23-OCT-2001
DEFINITION  Homo sapiens HIF1A mRNA for hypoxia-inducible factor 1 alpha
            variant, complete cds.
ACCESSION   AB073325
VERSION     AB073325.1  GI:16326343
KEYWORDS    .
SOURCE      Homo sapiens liver cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tanaka,S. and Sugimachi,K.
  TITLE     Hypoxia-inducible factor-1 alpha variant isolated from human liver
            tissue
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3551)
  AUTHORS   Tanaka,S. and Sugimachi,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-2001) Shinji Tanaka, Kyushu University, Graduate
            School of Medical Sciences, Department of Surgery and Science;
            3-1-1 Maidashi, Fukuoka, Fukuoka 812-8582, Japan
            (E-mail:shinjit@surg2.med.kyushu-u.ac.jp, Tel:81-92-642-5466,
            Fax:81-92-642-5482)
FEATURES             Location/Qualifiers
     source          1. .3551
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
     gene            1. .3551
                     /gene="HIF1A"
     CDS             29. .2236
                     /gene="HIF1A"
                     /note="HIF-1 alpha variant"
                     /codon_start=1
                     /product="hypoxia-inducible factor 1 alpha variant"
                     /protein_id="BAB70608.1"
                     /db_xref="GI:16326344"
                     /translation="MEGAGGANDKKKISSERRKEKSRDAARSRRSKESEVFYELAHQL
                     PLPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLDIEDDMKAQMNCFYLKALDGFVMV
                     LTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEMREMLTHRNGLVKKGKE
```

```
                     QNTQRSFFLRMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYDTNSNQPQCGYKKPPMT
                     CLVLICEPIPHPSNIEIPLDSKTFLSRHSLDMKFSYCDERITELMGYEPEELLGRSIY
                     EYYHALDSDHLTKTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQATVIYNTKNSQPQC
                     IVCVNYVVSGIIQHDLIFSLQQTECVLKPVESSDMKMTQLFTKVESEDTSSLFDKLKK
                     EPDALTLLAPAAGDTIISLDFGSNDTETDDQQLEEVPLYNDVMLPSPNEKLQNINLAM
                     SPLPTAETPKPLRSSADPALNQEVALKLEPNPESLELSFTMPQIQDQTPSPSDGSTRQ
                     SSPEPNSPSEYCFYVDSDMVNEFKLELVEKLFAEDTEAKNPFSTQDTDLDLEMLAPYI
                     PMDDDFQLRSFDQLSPLESSSASPESASPQSTVTVFQQTQIQEPTANATTTTATTDEL
                     KTVTKDRMEDIKILIASPSPTHIHKETTSATSSPYRDTQSRTASPNRAGKGVIEQTEK
                     SHPRSPNVLSVALSQRTTVPEEELNPKILALQNAQRKRKMEHDGSLFQAVGII"
BASE COUNT     1150 a    671 c    650 g   1080 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 9;  Length 3551;
  Best Local Similarity   96.0%;  Pred. No. 0.38;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCATGTGACCATTAGGAAATGAGAG 25
          ||||||||||||| |||||||||||||
Db    440 CCATGTGACCATGAGGAAATGAGAG 464


RESULT 12
HSU22431
LOCUS       HSU22431                3678 bp    mRNA    linear   PRI 28-JUN-1995
DEFINITION  Human hypoxia-inducible factor 1 alpha (HIF-1 alpha) mRNA, complete
            cds.
ACCESSION   U22431
VERSION     U22431.1  GI:881345
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3678)
  AUTHORS   Wang,G.L., Jiang,B.H., Rue,E.A. and Semenza,G.L.
  TITLE     Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS
            heterodimer regulated by cellular O2 tension
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92 (12), 5510-5514 (1995)
  MEDLINE   95296340
   PUBMED   7539918
REFERENCE   2  (bases 1 to 3678)
  AUTHORS   Wang,G.L., Jiang,B.-H., Rue,E.A. and Semenza,G.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-1995) Gregg L. Semenza, Center for Medical
            Genetics, The Johns Hopkins University School of Medicine, 600 N.
            Wolfe St., Baltimore, MD 21287-3914, USA
FEATURES             Location/Qualifiers
     source          1. .3678
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="Hep3B"
                     /cell_type="hepatoblastoma"
     gene            1. .3678
                     /gene="HIF-1 alpha"
     CDS             29. .2509
                     /gene="HIF-1 alpha"
                     /standard_name="hypoxia-inducible factor 1, alpha subunit"
                     /note="basic helix-loop-helix transcription factor"
                     /codon_start=1
                     /product="hypoxia-inducible factor 1 alpha"
                     /protein_id="AAC50152.1"
                     /db_xref="GI:881346"
                     /translation="MEGAGGANDKKKISSERRKEKSRDAARSRRSKESEVFYELAHQL
                     PLPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLDIEDDMKAQMNCFYLKALDGFVMV
                     LTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEMREMLTHRNGLVKKGKE
                     QNTQRSFFLRMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYDTNSNQPQCGYKKPPMT
                     CLVLICEPIPHPSNIEIPLDSKTFLSRHSLDMKFSYCDERITELMGYEPEELLGRSIY
                     EYYHALDSDHLTKTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQATVIYNTKNSQPQC
                     IVCVNYVVSGIIQHDLIFSLQQTECVLKPVESSDMKMTQLFTKVESEDTSSLFDKLKK
                     EPDALTLLAPAAGDTIISLDFGSNDTETDDQQLEEVPLYNDVMLPSPNEKLQNINLAM
                     SPLPTAETPKPLRSSADPALNQEVALKLEPNPESLELSFTMPQIQDQTPSPSDGSTRQ
                     SSPEPNSPSEYCFYVDSDMVNEFKLELVEKLFAEDTEAKNPFSTQDTDLDLEMLAPYI
                     PMDDDFQLRSFDQLSPLESSSASPESASPQSTVTVFQQTQIQEPTANATTTTATTDEL
                     KTVTKDRMEDIKILIASPSPTHIHKETTSATSSPYRDTQSRTASPNRAGKGVIEQTEK
                     SHPRSPNVLSVALSQRTTVPEEELNPKILALQNAQRKRKMEHDGSLFQAVGIGTLLQQ
                     PDDHAATTSLSWKRVKGCKSSEQNGMEQKTIILIPSDLACRLLGQSMDESGLPQLTSY
                     DCEVNAPIQGSRNLLQGEELLRALDQVN"
     polyA_site      3678
                     /gene="HIF-1 alpha"
                     /note="42 A nucleotides"
BASE COUNT     1197 a    695 c    675 g   1111 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 9;  Length 3678;
  Best Local Similarity   96.0%;  Pred. No. 0.38;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCATGTGACCATTAGGAAATGAGAG 25
          ||||||||||||| |||||||||||||
Db    440 CCATGTGACCATGAGGAAATGAGAG 464


RESULT 13
AF057308
LOCUS       AF057308                3718 bp    mRNA    linear   ROD 04-SEP-2001
DEFINITION  Rattus norvegicus hypoxia-inducible factor-1 alpha (Hif1a) mRNA,
            complete cds.
ACCESSION   AF057308
VERSION     AF057308.1  GI:4580532
KEYWORDS    .
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 3718)
  AUTHORS   Zou,A.P., Yang,Z.Z., Li,P.L. and Cowley AW,J.R.
  TITLE     Oxygen-dependent expression of hypoxia-inducible factor-1alpha in
            renal medullary cells of rats
  JOURNAL   Physiol. Genomics (Online) 6 (3), 159-168 (2001)
  MEDLINE   21417706
   PUBMED   11526200
REFERENCE   2  (bases 1 to 3718)
  AUTHORS   Zou,A.P., Su,N., Park,F., Li,P.L. and Cowley,A.W. Jr.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-1998) Departments of Physiology and Pharmacology,
            Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee,
            WI 53226, USA
FEATURES             Location/Qualifiers
     source          1. .3718
                     /organism="Rattus norvegicus"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10116"
                     /sex="male"
                     /tissue_type="kidney"
                     /dev_stage="adult"
     gene            1. .3718
                     /gene="Hif1a"
     CDS             24. .2495
                     /gene="Hif1a"
                     /note="transcription factor; RH1F-1"
                     /codon_start=1
                     /product="hypoxia-inducible factor-1 alpha"
                     /protein_id="AAD24413.1"
                     /db_xref="GI:4580533"
                     /translation="MEGAGGENEKKNRMSSERRKEKSRDAARSRRSKESEVFYELAHQ
                     LPLPHNVSSHLDKASVMRLTISYLRVRKLLGAGDLDIEDEMKAQMNCFYLKALDGFVM
                     VLTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEMREMLTHRNGPVRKGK
                     EQNTQRSFFLRMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYDTSSNQPQCGYKKPPM
                     TCLVLICEPIPHPSNIEIPLDSKTFLSRHSLDMKFSYCDERITELMGYEPEELLGRSI
                     YEYYHALDSDHLTKTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQATVIYNTKNSQPQ
                     CIVCVNYVVSGIIQHDLIFSLQQTESVLKPVESSDMKMTQLFTKVESEDTSCLFDKLK
                     KEPDALTLLAPAAGDTIISLDFGSDDTETEDQQLEDVPLYNDVMFPSSNEKLNINLAM
                     SPLPASETPKPLRSSADPALNQEVALKLESSPESLGLSFTMPQIQDQPASPSDGSTRQ
                     SSPEPNSPSEYCFDVDSDMVNVFKLELVEKLFAEDTEAKNPFSAQDTDLDLEMLAPYI
```

PMDDDFQLRSFDQLSPLESNSPSPPSVSTVTGFQQTQLQKPTITVTATATATTDESKA
VTKDNIEDIKILIASPPSTQVPQEMTTAKASAYSGTHSRTASPDRAGKRVIEKTDKAH
PRSLNLSVTLNQRNTVPEEELNPKTIALQNAQRKRKMEHDGSLFQAAGIGTLLQQPGD
RAPTMSLSWKRVKGYISSEQDGMEQKTIFLIPSDLACRLLGQSMDESGLPQLTSYDCE
VNAPIQGSRNLLQGEELLRALDQVN"

BASE COUNT 1125 a 733 c 762 g 1098 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 10; Length 3718;
Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||||| |||||||||||
Db 438 CCATGTGACCATGAGGAAATGAGAG 462

RESULT 14
AX306008
LOCUS AX306008 3746 bp DNA linear PAT 11-DEC-2001
DEFINITION Sequence 759 from Patent WO0188188.
ACCESSION AX306008
VERSION AX306008.1 GI:17645352
KEYWORDS .
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Ishikawa,K., Asai,S., Takahashi,Y., Nagata,T. and Ishii,Y.
TITLE Method for examining ischemic conditions
JOURNAL Patent: WO 0188188-A 759 22-NOV-2001;
School Juridical Person Nihon University (JP)
FEATURES Location/Qualifiers
source 1. .3746
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 1124 a 740 c 740 g 1142 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 6; Length 3746;
Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||||| |||||||||||
Db 501 CCATGTGACCATGAGGAAATGAGAG 525

RESULT 15
MMHIF1ALP
LOCUS MMHIF1ALP 3746 bp mRNA linear ROD 16-JUL-1996
DEFINITION M.musculus mRNA for hypoxia-inducible factor I alpha.
ACCESSION X95580
VERSION X95580.1 GI:1430864
KEYWORDS HIF-1alpha; hypoxia-inducible factor one alpha.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3746)
AUTHORS Wenger,R.H.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-1996) R.H. Wenger, Institute of Physiology,
University of Zuerich-Irchel, Winterthurerstrasse 190, CH-8057
Zuerich, SWITZERLAND
REMARK Revised by author 16-JUL-96
REFERENCE 2 (bases 1 to 3746)
AUTHORS Wenger,R.H., Rolfs,A., Marti,H.H., Guenet,J.L. and Gassmann,M.
TITLE Nucleotide sequence, chromosomal assignment and mRNA expression of
mouse hypoxia-inducible factor-1 alpha
JOURNAL Biochem. Biophys. Res. Commun. 223 (1), 54-59 (1996)
MEDLINE 96254028
PUBMED 8660378
COMMENT On Jul 17, 1996 this sequence version replaced gi:1419626.
Related sequence X95002 and U59496.
FEATURES Location/Qualifiers
source 1. .3746
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="Hepalc1c7"
/cell_type="hepatoma"
gene 126. .2558
/gene="HIF-1alpha"
CDS 126. .2558
/gene="HIF-1alpha"
/codon_start=1
/product="hypoxia-inducible factor one alpha"
/protein_id="CAA64833.1"
/db_xref="GI:4379202"
/db_xref="MGD:MGI:106918"
/db_xref="SWISS-PROT:Q61221"
/translation="MSSERRKEKSRDAARSRRSKESEVFYELAHQLPLPHNVSSHLDK
ASVMRLTISYLRVRKLLDAGGLDSEDEMKAQMDCFYLKALDGFVMVLTDDGDMVYISD
NVNKYMGLTQFELAGHSVFDFTHPCDHEEMREMLTHRNGPVRKGKELNTQRSFFLRMK
CTLTSRGRTMNIKSATWKVLHCTGHIHVYDTNSNQPQCGYKKPPMTCLVLICEPIPHP
SNIEIPLDSKTFLSRHSLDMKFSYCDERITELMGYEPEELLGRSIYEYYHALDSDHLT
KTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQATVIYNTKNSQPQCIVCVNYVVSGII
QHDLIFSLQQTESVLKPVESSDMKMTQLFTKVESEDTSCLFDKLKKEPDALTLLAPAA
GDTIISLDFGSDDTETEDQQLEDVPLYNDVMFPSSNEKLNINLAMSPLPSSETPKPLR
SSADPALNQEVALKLESSPESLGLSFTMPQIQDQPASPSDGSTRQSSPEPNSPSEYCF
DVDSDMVNVFKLELVEKLFAEDTEAKNPFSTQDTDLDLEMLAPYIPMDDDFQLRSFDQ
LSPLESNSPSPPSMSTVTGFQQTQLQKPTITATATTTATTDESKTETKDNKEDIKILI
ASPSSTQVPQETTTAKASAYSGTHSRTASPDRAGKRVIEQTDKAHPRSLNLSATLNQR
NTVPEEELNPKTIASQNAQRKRKMEHDGSLFQAAGIGTLLQQPGDCAPTMSLSWKRVK
GFISSEQNGTEQKTIILIPSDLACRLLGQSMDVSGLPQLTSYDCEVNAPIQGSRNLLQ
GEELLRALDQVN"
BASE COUNT 1124 a 740 c 740 g 1142 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 10; Length 3746;
Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||||| |||||||||||
Db 501 CCATGTGACCATGAGGAAATGAGAG 525

Search completed: March 17, 2003, 11:24:12
Job time : 236.108 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:42:12 ; Search time 143.28 Seconds
(without alignments)
392.938 Million cell updates/sec

Title: US-09-836-439-2
Perfect score: 25
Sequence: 1 ccatgtgaccattaggaaatgagag 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	23.4	93.6	2481	24	ABK11552	Human cDNA encodin
	2	23.4	93.6	2481	24	ABL91695	Human polynucleoti
	3	23.4	93.6	2528	22	AAS14154	Human HIF-1 alpha
	4	23.4	93.6	3678	19	AAV63210	Nucleic acid seque
	5	23.4	93.6	3678	24	AAD38995	Human HIF-1alpha c
	6	23.4	93.6	3678	24	ABK84267	Human cDNA differe
	7	23.4	93.6	3736	18	AAT45937	Human hypoxia indu
	8	23.4	93.6	3736	21	AAZ99537	DNA encoding a wil
	9	23.4	93.6	3746	24	ABI99710	Mouse ischaemic co
	10	23.4	93.6	3927	24	AAS61690	Lung small cell ca
	11	23.4	93.6	3933	20	AAX58980	Human transcriptio
	12	23.4	93.6	4162	23	AAS85058	DNA encoding novel
	13	23.4	93.6	10355	22	AAS14156	pSMART CMV-HIF DNA
c	14	23.4	93.6	27884	22	AAK77781	Human immune/haema
c	15	18.2	72.8	367	22	AAH93977	Human foetal cDNA,
	16	18.2	72.8	589	24	ABK62425	Rat sequence diffe
	17	18.2	72.8	1070	21	AAC75526	Human ORFX ORF1081
	18	18.2	72.8	1070	24	ABN77923	Human ORF2870 cDNA
c	19	17.6	70.4	353	21	AAC28821	Human secreted pro
	20	17.6	70.4	439	23	ABV13952	Human prostate exp
	21	17.6	70.4	477	23	ABV35061	Human prostate exp
	22	17.6	70.4	477	23	ABV43908	Human prostate exp
	23	17.6	70.4	480	23	ABV04783	Human prostate exp
	24	17.6	70.4	813	21	AAA44916	Human secreted exp
	25	17.6	70.4	829	22	AAH06353	Human cDNA clone (
c	26	17.6	70.4	1174	21	AAF15982	Human prostate can
c	27	17.6	70.4	1174	22	AAH33713	Human colon cancer
c	28	17.6	70.4	1707	23	AAS81186	DNA encoding novel
c	29	17.6	70.4	2158	23	AAS78604	DNA encoding novel
c	30	17.6	70.4	2159	23	AAS71921	DNA encoding novel
	31	17.6	70.4	2280	22	AAH18074	Human cDNA sequenc
	32	17.6	70.4	2987	22	AAI93889	Human stomach canc
	33	17.6	70.4	2987	22	AAH16184	Human cDNA sequenc
c	34	17.6	70.4	3183	23	AAS71919	DNA encoding novel
c	35	17.6	70.4	3183	23	AAS84255	DNA encoding novel
c	36	17.6	70.4	3193	23	AAS78605	DNA encoding novel
	37	17.6	70.4	3625	22	AAH46859	Human serine/threo
	38	17.6	70.4	3625	22	AAH78738	Human HSQFV91 seri
	39	17.6	70.4	3708	24	ABA05449	Human transcriptio
c	40	17.6	70.4	4172	23	ABV23038	Human prostate exp
c	41	17.6	70.4	4172	23	ABV28874	Human prostate exp
	42	17.6	70.4	7545	22	ABA09525	Human secreted pro
	43	17.6	70.4	7545	22	AAK53401	Human polynucleoti
	44	17.6	70.4	7545	22	AAK53402	Human polynucleoti
c	45	17.6	70.4	21469	22	AAK89568	Human digestive sy

ALIGNMENTS

RESULT 1
ABK11552
ID ABK11552 standard; cDNA; 2481 BP.
XX
AC ABK11552;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human cDNA encoding hypoxia-inducible factor-1 alpha, HIF-1.
XX
KW Human; ss; gene; HIF-1; hypoxia-inducible factor-1;
KW transactivation domain; N-TAD; C-TAD; ischaemia; brain infarction;
KW circulatory disorder; cancer; hypertension; demyelinating disorder;
KW angiogenesis; sarcoidosis; hepatitis-caused inflammation;
KW chronic ulceration; neovascularisation; arterial hypervascularisation;
KW bullous skin disease; vasculitis; dermatomyositis; polymyositis;
KW rheumatoid arthritis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..2481
FT /*tag= a
FT /product= "HIF-1"
XX
PN WO200212326-A2.
XX
PD 14-FEB-2002.
XX
PF 07-AUG-2001; 2001WO-IB01775.
XX
PR 07-AUG-2000; 2000US-223480P.

XX
PA (ASPE-) ASPERA PHARM AB.
XX
PI Poellinger L, Pereira T, Ruas J;
XX
DR WPI; 2002-257466/30.
DR P-PSDB; AAU77602.
XX
PT New polypeptides comprising hypoxia-inducible factor-1 with alterations
PT of the transactivation domain, useful treating ischaemic conditions,
PT e.g. brain infarction, heart infarction or circulatory disorder -
XX
PS Disclosure; Page 73-76; 80pp; English.
XX
CC The invention relates to a polypeptide comprising hypoxia-inducible
CC factor-1 (HIF-1) with alterations of the transactivation domain (N-TAD or
CC C-TAD). Also included are nucleic acids encoding the altered proteins, a
CC vector comprising the nucleic acid, a host cell transformed with the
CC vector, methods for producing the protein or its functional fragment
CC or an isolated degradation box, a method of screening for an agent that
CC modulates N-TAD function and antagonists, agonists, modulators and
CC HIF-1 peptide fragments useful for modulating HIF-1 function or the
CC function of proteins that interact with it. The isolated polypeptides and
CC their fragments with altered residues are useful in methods for treating
CC diseases. The disease is an ischaemic condition, e.g. brain infarction,
CC heart infarction or circulatory disorder. The disease may also be cancer,
CC hypertension, demyelinating disorders, diffuse proliferative
CC glomerulonephritis, toxoplasmosis caused retinochorioiditis, HIV (human
CC immunodeficiency virus) caused Tat angiogenesis, HIV-caused Kaposi's
CC sarcoma, hepatitis-caused inflammation, hepatitis-caused angiogenesis,
CC chronic ulceration, proliferative retinopathy, retina haemangioblastomas,
CC neovascularisation, arterial hypervascularisation, sarcoidosis, bullous
CC skin disease, vasculitis with angiogenesis, dermatomyositis with
CC angiogenesis, polymyositis with angiogenesis, rheumatoid arthritis,
CC juvenile osteoarthritis, polyarthritis, aneurysm or atheroma. The
CC present sequence encodes HIF-1.
XX
SQ Sequence 2481 BP; 829 A; 512 C; 500 G; 640 T; 0 other;

Query Match 93.6%; Score 23.4; DB 24; Length 2481;
Best Local Similarity 96.0%; Pred. No. 0.19;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||||| ||||||||||||
Db 412 CCATGTGACCATGAGGAAATGAGAG 436

RESULT 2
ABL91695
ID ABL91695 standard; DNA; 2481 BP.
XX
AC ABL91695;
XX
DT 28-MAY-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 38.
XX
KW Human; HIV; HCV; gene expression; oligoribonucleotide; tumour; pathogen;
KW Plasmodium; virus; viroid; cytokine; prion; antisense oligonucleotide;
KW cytostatic; virucide; protozoacide; antibacterial; ds.
XX
OS Homo sapiens.
XX
PN DE10100586-C1.
XX
PD 11-APR-2002.
XX
PF 09-JAN-2001; 2001DE-1000586.
XX
PR 09-JAN-2001; 2001DE-1000586.
XX
PA (RIBO-) RIBOPHARMA AG.
XX
PI Kreutzer R, Limmer S, Rost S, Hadwiger P;
XX
DR WPI; 2002-270454/32.
XX
PT Inhibiting gene expression in cells, useful for e.g. treating tumors,
PT by introducing double-stranded complementary oligoRNA having unpaired
PT terminal bases -
XX
PS Claim 13; Page 33; 104pp; German.
XX
CC The invention relates to a method for inhibiting expression of a target
CC gene (ABL91658-ABL91797) in a cell by introducing at least one
CC oligoribonucleotide that has a double-stranded structure consisting of at
CC most 49 sequential nucleotide pairs, with at least part of one strand
CC complementary with the target gene and has at least one end a
CC single-stranded segment of 1-4 nt. The method provides
CC oligoribonucleotides for antisense inhibition of gene expression useful
CC e.g. for treating tumours but the oligoribonucleotides may also be
CC directed against genes present in pathogens (e.g. Plasmodium or
CC viruses/viroids, pathogenic on humans, animals or plants) or against
CC cytokine, Id, developmental or prion genes. The method provides more
CC effective inhibition of gene expression than use of known
CC oligonucleotides, probably because the unpaired overhang increases
CC stability and thus intracellular concentration.
XX
SQ Sequence 2481 BP; 829 A; 512 C; 500 G; 640 T; 0 other;

Query Match 93.6%; Score 23.4; DB 24; Length 2481;
Best Local Similarity 96.0%; Pred. No. 0.19;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||||| ||||||||||||
Db 412 CCATGTGACCATGAGGAAATGAGAG 436

RESULT 3
AAS14154
ID AAS14154 standard; cDNA; 2528 BP.
XX
AC AAS14154;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human HIF-1 alpha DNA used in identification of hypoxia regulated genes.
XX
KW Differential expression; polymorphism; biological pathway; IRES; GFP; ss;
KW internal ribosome entry site; green fluorescent protein; HIF-1 alpha;
KW hypoxia inducible factor 1 alpha; hypoxia regulated gene; macrophage;
KW human.
XX
OS Homo sapiens.
XX
PN WO200162965-A2.
XX
PD 30-AUG-2001.
XX
PF 22-FEB-2001; 2001WO-GB00758.
XX
PR 22-FEB-2000; 2000GB-0004197.
PR 28-JUL-2000; 2000GB-0018679.
XX
PA (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
PI Kingsman AJ;
XX
DR WPI; 2001-589807/66.
XX
PT Screening a genetic element involved in a cellular process, comprises
PT comparing gene expressions in a cell, and a second cell that has
PT altered levels of genes used in the process, and detecting an element
PT with varied expression -

XX
PS Example 2; Page 96-97; 103pp; English.
XX
CC The invention relates to differential expression screening to identify a
CC genetic element involved in a cellular process. The method involves
CC comparing gene expressions in two cells, where one cell has altered
CC levels of a biological molecule, and identifying the genetic element
CC whose expression differs, by comparing expression under different
CC environmental conditions. The method is useful for identifying mutations
CC and polymorphisms that affect the biological response to a particular
CC cellular process. The method also allows the molecular dissection of
CC biological pathways by altering a particular pathway under study. By
CC introducing a heterologous nucleic acid into a cell population, the level
CC of a particular molecule can be influenced, allowing a pathway to be
CC dissected into its precise molecular components. The main use of the
CC method is to compare gene expression in cells from a diseased patient and
CC from a normal patient. This sequence represents a human hypoxia inducible
CC factor 1 alpha (HIF-1 alpha) which is inserted into an AdApt ires-GFP
CC plasmid for use in identification of hypoxia-regulated genes in
CC macrophages.
XX
SQ Sequence 2528 BP; 839 A; 528 C; 513 G; 648 T; 0 other;

Query Match 93.6%; Score 23.4; DB 22; Length 2528;
Best Local Similarity 96.0%; Pred. No. 0.19;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||||| ||||||||||||
Db 439 CCATGTGACCATGAGGAAATGAGAG 463

RESULT 4
AAV63210
ID AAV63210 standard; DNA; 3678 BP.
XX
AC AAV63210;
XX
DT 14-JAN-1999 (first entry)
XX
DE Nucleic acid sequence of human hypoxia inducible factor 1 alpha.
XX
KW Human transforming growth factor-beta 3; TGF-beta3; oxygen tension;
KW trophoblast invasion regulation; inhibitor; HIF-1 alpha;
KW TGF-beta family cytokine receptor; hypoxia inducible factor 1 alpha;
KW preeclempsia; pregnanancy; choriocarcinoma; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 29..2509
FT /*tag= a
FT /product= HIF-1 alpha
XX
PN WO9840747-A1.
XX
PD 17-SEP-1998.
XX
PF 05-MAR-1998; 98WO-CA00180.
XX
PR 07-MAR-1997; 97US-0039919.
XX
PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
PA (MOUN) MOUNT SINAI HOSPITAL CORP.
XX
PI Caniggia I, Lye S, Post M;
XX
DR WPI; 1998-520837/44.
DR P-PSDB; AAW80418.
XX
PT Regulation of trophoblast invasion - by, e.g. transforming growth
PT factor-beta3 inhibitor, useful for detecting or treating
PT preeclempsia in pregnant women
XX
PS Disclosure; Fig 2; 59pp; English.
XX
CC The present sequence encodes human hypoxia inducible factor 1 alpha
CC (HIF-1 alpha). The specification describes a composition for regulating
CC trophoblast invasion which comprises an inhibitor of transforming growth
CC factor-beta 3 (TGF-beta3), TGF-beta family cytokine receptors, HIF-1
CC alpha or oxygen tension. The composition is used in methods of
CC diagnosing, monitoring, preventing or treating conditions requiring
CC regulation of trophoblast invasion, especially preeclempsia in pregnant
CC women or choriocarcinomas.
XX
SQ Sequence 3678 BP; 1197 A; 695 C; 675 G; 1111 T; 0 other;

Query Match 93.6%; Score 23.4; DB 19; Length 3678;
Best Local Similarity 96.0%; Pred. No. 0.2;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||||| ||||||||||||
Db 440 CCATGTGACCATGAGGAAATGAGAG 464

RESULT 5
AAD38995
ID AAD38995 standard; cDNA; 3678 BP.
XX
AC AAD38995;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human HIF-1alpha cDNA.
XX
KW Human; haematologic malignancy; multidrug resistance; MDR; SUMO-1;
KW hypoxia inducible factor-1; small ubiquitin-like-modifier; HIF-1;
KW lymphoid disorder; chronic lymphoproliferative disorder; lymphoma;
KW myeloid disorder; lymphocytic leukaemia; thrombocythaemia; myeloma;
KW angiogenic myeloid metaplasia; myeloid leukaemia; gene therapy;
KW polycythaemia vera; hypoxia responsive element; HRE; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 29..2509
FT /*tag= a
FT /product= "Human HIF-1alpha protein"
XX
PN WO200234291-A2.
XX
PD 02-MAY-2002.
XX
PF 25-OCT-2001; 2001WO-US49856.
XX
PR 26-OCT-2000; 2000US-243542P.
XX
PA (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
XX
PI Colgan SP;
XX
DR WPI; 2002-471427/50.
DR P-PSDB; AAE24212.
XX
PT Treating a subject (at risk of) having a hematologic malignancy or
PT multidrug resistance, e.g. lymphoma or myeloma, by administering
PT hypoxia inducible factor 1 binding molecules or small
PT ubiquitin-like-modifier-1 binding molecules -
XX
PS Disclosure; Page 61-63; 92pp; English.
XX
CC The invention relates to a method of treating a subject having or at
CC risk of developing a haematologic malignancy or multidrug resistance
CC (MDR). The method involves administering hypoxia inducible factor-1
CC (HIF-1) binding molecules or small ubiquitin-like-modifier (SUMO)-1

```
CC   binding molecules or HIF-1-SUMO-1 complex modulators. mdr1-hypoxia
CC   responsive element (HRE) binding molecules or antisense nucleic
CC   acid molecules and SUMO-1 binding molecules or antisense molecules
CC   are useful for treating a subject having or at risk of developing
CC   haematologic malignancy or MDR (e.g. a lymphoid disorder or a myeloid
CC   disorder). The lymphoid disorders include lymphocytic leukaemia or
CC   chronic lymphoproliferative disorders e.g. lymphoma, myeloma or chronic
CC   lymphoid leukaemia. The myeloid disorders include chronic or acute
CC   myeloid leukaemia, e.g. angiogenic myeloid metaplasia, essential
CC   thrombocythaemia or polycythaemia vera. The invention is used in gene
CC   therapy. The present sequence is human HIF-1alpha cDNA.
XX
SQ   Sequence 3678 BP; 1197 A; 695 C; 675 G; 1111 T; 0 other;

  Query Match             93.6%;  Score 23.4;  DB 24;  Length 3678;
  Best Local Similarity   96.0%;  Pred. No. 0.2;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||||  ||||||||||||
Db     440 CCATGTGACCATGAGGAAATGAGAG 464


RESULT 6
ABK84267
ID   ABK84267 standard; cDNA; 3678 BP.
XX
AC   ABK84267;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #838.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US30821.
XX
PR   03-OCT-2000; 2000US-237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression
PT   of genes associated with granulocyte activation, which serves as
PT   diagnostic markers that is useful for monitoring disease states and
PT   drug toxicity  -
XX
PS   Claim 1; SEQ ID No 838; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing
CC   the expression level to an expression level in an unactivated
CC   GC, where differential expression of Gs is indicative of GCA.
CC   Also included are modulating (M2) GA by contacting GC with an agent
CC   that alters the expression of at least one gene in Gs; (2) screening (M3)
CC   for an agent capable of modulating GCA or an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease using the
CC   gene expression profile; (3) detecting (M4) an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease, by detecting the
CC   level of expression in a sample of the tissue of gene(s) from Gs, where
CC   the level of expression of the gene is indicative of inflammation;
CC   (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC   an allergic response in a subject, exposure of a subject to a pathogen
CC   or sterile inflammatory disease, by contacting a tissue having
CC   inflammation with an agent that modulates the expression of gene(s)
CC   from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC   modulating GA; M3 is useful for screening an agent capable of modulating
CC   GCA preferably in an inflammation in a tissue; M4 is useful for
CC   detecting an inflammation (especially chronic) in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC   glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC   reperfusion injury, ARDS, adult respiratory distress syndrome,
CC   inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC   periodontal disease; also bacterial infection, viral infection,
CC   parasitic infection, protozoal infection, fungal infection and M5 is
CC   useful for treating one of the above conditions. The present
CC   sequence represents a gene differentially expressed in granulocytes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3678 BP; 1197 A; 695 C; 675 G; 1111 T; 0 other;

  Query Match             93.6%;  Score 23.4;  DB 24;  Length 3678;
  Best Local Similarity   96.0%;  Pred. No. 0.2;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||||  ||||||||||||
Db     440 CCATGTGACCATGAGGAAATGAGAG 464


RESULT 7
AAT45937
ID   AAT45937 standard; DNA; 3736 BP.
XX
AC   AAT45937;
XX
DT   19-MAR-1997  (first entry)
XX
DE   Human hypoxia inducible factor-1 alpha cDNA.
XX
KW   Hypoxia inducible factor-1 alpha; HIF-1; tissue damage;
KW   atherosclerosis; cerebral artery disease; gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_signal      23..32
FT                    /*tag=  a
FT                    /function= Kozak sequence
FT   CDS              29..2509
FT                    /*tag=  b
FT   polyA_signal     3674..3679
FT                    /*tag=  c
XX
PN   WO9639426-A1.
XX
PD   12-DEC-1996.
XX
PF   06-JUN-1996;   96WO-US10251.
XX
PR   06-JUN-1995;   95US-0480473.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS SCHOOL MED.
```

```
XX
PI   Semenza GL;
XX
DR   WPI; 1997-043061/04.
DR   P-PSDB; AAW06557.
XX
PT   DNA encoding human hypoxia-inducible factor 1 alpha - useful for
PT   enhancing expression of structural gene and treatment of
PT   hypoxia-related tissue damage
XX
PS   Claim 2; Page 49-53; 95pp; English.
XX
CC   A DNA sequence (AAT45937) codes for the alpha subunit (AAW06557) of
CC   human hypoxia inducible factor-1 (HIF-1), a DNA binding protein
CC   that binds to the enhancer region of e.g. erythropoietin and
CC   vascular endothelial growth factor (VEGF) genes.  The DNA sequence
CC   is a composite of 3 clones obtd. by screening an Hep3B library and
CC   by database analysis.  HIF-1 alpha nucleic acids can be used to
CC   prepare recombinant HIF-1 alpha in transformed host cells, as
CC   probes, and in the gene therapy of HIF-1-mediated or hypoxia-related
CC   disorders such as atherosclerotic coronary or cerebral artery
CC   disease; antisense sequences inhibit HIF-1 expression e.g. to treat
CC   tumour proliferation mediated by VEGF-induced angiogenesis.
XX
SQ   Sequence 3736 BP; 1255 A; 695 C; 675 G; 1111 T; 0 other;

  Query Match             93.6%;  Score 23.4;  DB 18;  Length 3736;
  Best Local Similarity   96.0%;  Pred. No. 0.2;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||| ||||||||||||
Db     440 CCATGTGACCATGAGGAAATGAGAG 464


RESULT 8
AAZ99537
ID   AAZ99537 standard; DNA; 3736 BP.
XX
AC   AAZ99537;
XX
DT   03-JUL-2000  (first entry)
XX
DE   DNA encoding a wild type human hypoxia inducible factor-1 alpha.
XX
KW   Human; hypoxia-inducible factor 1 alpha; HIF-1alpha; variant;
KW   hypoxia inducible gene; hypoxia inducible factor; hypoxia;
KW   ischemia related damage; angiogenesis; coronary artery disease;
KW   ischemic tissue damage; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             29..2509
FT                   /*tag= a
FT                   /product= "hypoxia inducible factor-1 alpha"
XX
PN   WO200010578-A1.
XX
PD   02-MAR-2000.
XX
PF   25-AUG-1999;   99WO-US19416.
XX
PR   25-AUG-1998;   98US-0148547.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI   Semenza GL;
XX
DR   WPI; 2000-246493/21.
DR   P-PSDB; AAY69407.
XX
PT   Variant forms of hypoxia-inducible factor (HIF)-1 alpha, useful for
PT   treating hypoxia or ischemia-related tissue damage -
XX
PS   Disclosure; Page 80-89; 96pp; English.
XX
CC   The present sequence encodes a wild type human hypoxia-inducible factor
CC   (HIF)-1 alpha. The specification describes HIF-1alpha variants that are
CC   stable under hypoxic and non-hypoxic conditions. The variants comprises
CC   amino acid residues 1-391 and 521-826, 549-826, 576-826, 429-826,
CC   469-826, 494-826, 508-826, 512-826 or 517-826 of the wild type human
CC   HIF-1alpha polypeptide, in which residues 551 and 552 are not serine
CC   and threonine, respectively. The HIF-1alpha variant polynucleotide
CC   sequences are useful for increasing expression of a hypoxia inducible
CC   gene in a cell. They is also useful for providing constitutive
CC   expression of a hypoxia inducible factor in a cell, and for reducing
CC   or preventing hypoxia or ischemia related damage. The variant
CC   HIF-1alpha polypeptides are useful for providing prophylactic therapy
CC   for inducing the level of angiogenesis in tissues of patients at risk
CC   of coronary artery disease or ischemic tissue damage.
XX
SQ   Sequence 3736 BP; 1255 A; 695 C; 675 G; 1111 T; 0 other;

  Query Match             93.6%;  Score 23.4;  DB 21;  Length 3736;
  Best Local Similarity   96.0%;  Pred. No. 0.2;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||| ||||||||||||
Db     440 CCATGTGACCATGAGGAAATGAGAG 464


RESULT 9
ABI99710
ID   ABI99710 standard; cDNA; 3746 BP.
XX
AC   ABI99710;
XX
DT   07-MAR-2002  (first entry)
XX
DE   Mouse ischaemic condition related cDNA sequence SEQ ID NO:759.
XX
KW   Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW   vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
OS   Mus musculus.
XX
PN   WO200188188-A2.
XX
PD   22-NOV-2001.
XX
PF   18-MAY-2001; 2001WO-JP04192.
XX
PR   18-MAY-2000; 2000JP-0145977.
XX
PA   (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
PI   Ishikawa K,  Asai S,  Takahashi Y,  Nagata T,  Ishii Y;
XX
DR   WPI; 2002-034733/04.
DR   P-PSDB; ABB57270.
XX
PT   Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT   expression levels of particular genes defined in the specification or
PT   by determining the expression profile of a gene group comprising these
PT   genes -
XX
PS   Claim 2; Page 1863-1869; 2690pp; English.
XX
CC   The present invention describes a method for examining ischaemic
CC   conditions, comprising measuring the expression levels of particular
CC   genes (I) in a test sample or determining the expression profile of a
CC   gene group in the sample comprising genes selected from (I). The method
CC   is useful for examining the ischaemic condition (e.g. compressive
CC   ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
```

```
CC   expression levels of particular genes (ABI99202 to ABI99912, encoding
CC   the protein sequences in ABB57020 to ABB57374) or by determining the
CC   expression profile of a gene group comprising these genes. The
CC   expression levels or expression profiles produced by these genes are
CC   used as an indicator when screening for ischaemic condition-improving
CC   drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
CC   represent PCR primers for a mouse ischaemic condition related sequence,
CC   which are used in the exemplification of the present invention.
XX
SQ   Sequence 3746 BP; 1124 A; 740 C; 740 G; 1142 T; 0 other;

  Query Match             93.6%;  Score 23.4;  DB 24;  Length 3746;
  Best Local Similarity   96.0%;  Pred. No. 0.2;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||| ||||||||||||
Db     501 CCATGTGACCATGAGGAAATGAGAG 525


RESULT 10
AAS61690
ID   AAS61690 standard; cDNA; 3927 BP.
XX
AC   AAS61690;
XX
DT   29-JAN-2002  (first entry)
XX
DE   Lung small cell carcinoma antigen, cDNA #231.
XX
KW   Human; cytostatic; antitumour; lung small cell cancer antigen;
KW   tumour; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200177168-A2.
XX
PD   18-OCT-2001.
XX
PF   11-APR-2001; 2001WO-US11859.
XX
PR   11-APR-2000; 2000US-196780P.
PR   21-JUN-2000; 2000US-213361P.
PR   01-SEP-2000; 2000US-229763P.
PR   05-SEP-2000; 2000US-230629P.
PR   14-SEP-2000; 2000US-232565P.
PR   19-DEC-2000; 2000US-257037P.
PR   08-JAN-2001; 2001US-260796P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Lodes MJ,  Wang T,  Mohamath R,  Indirias CY;
XX
DR   WPI; 2002-010896/01.
DR   P-PSDB; AAU69409.
XX
PT   Lung tumour polynucleotide and polypeptides useful in therapy and
PT   diagnosis of cancer especially lung cancer  -
XX
PS   Claim 1; Page 201-202; 295pp; English.
XX
CC   The invention relates to novel isolated lung small cell cancer antigen
CC   polynucleotides (I) and polypeptides (II) used in a method of detecting
CC   cancer in a patient. The method is optionally performed by
CC   utilising oligonucleotides (III), where the biological sample
CC   from the patient is contacted with (III), detecting the amount of
CC   polynucleotide hybridised to (III) in the sample and comparing the
CC   amount of polynucleotide to a predetermined cut-off value and thereby
CC   determining cancer in a patient. (I), (II) or antigen-presenting cells
CC   expressing (II) is useful for stimulating and/or expanding T cells
CC   specific for a tumour protein. The method comprises contacting T cells
CC   with one of the components under conditions to permit the stimulation
CC   and/or expansion of the cells. A composition comprising (I) is useful for
CC   stimulating an immune response in a patient and for inhibiting the
CC   development of a cancer especially lung cancer in a patient. An
CC   isolated T cell population is useful for removing tumour cells from the
CC   biological sample and for inhibiting the development of cancer in a
CC   patient. AAS61460-AAS61874 represent novel human lung small cell
CC   cancer antigen coding sequences of the invention.
XX
SQ   Sequence 3927 BP; 1241 A; 782 C; 748 G; 1156 T; 0 other;

  Query Match             93.6%;  Score 23.4;  DB 24;  Length 3927;
  Best Local Similarity   96.0%;  Pred. No. 0.21;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||| ||||||||||||
Db     670 CCATGTGACCATGAGGAAATGAGAG 694


RESULT 11
AAX58980
ID   AAX58980 standard; cDNA; 3933 BP.
XX
AC   AAX58980;
XX
DT   23-AUG-1999  (first entry)
XX
DE   Human transcription regulator MOP1 cDNA.
XX
KW   MOP1; member of the PAS superfamily; bHLH-PAS; human;
KW   transcription regulator; hypoxia inducible factor 1 alpha; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             265..2745
FT                   /*tag=  a
XX
PN   WO9928464-A2.
XX
PD   10-JUN-1999.
XX
PF   27-NOV-1998;   98WO-US25314.
XX
PR   28-NOV-1997;   97US-0066863.
XX
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI   Bradfield CA,  Gu YZ,  Hogenesch JB;
XX
DR   WPI; 1999-371120/31.
DR   P-PSDB; AAY06289.
XX
PT   Developmental signal transduction associated proteins
XX
PS   Example 1; Page 93-94; 106pp; English.
XX
CC   This is the nucleotide sequence of a cDNA encoding MOP1 (see
CC   AAY06289), a member of the PAS superfamily, where PAS stands for
CC   PER/ARNT/SIM domains.  The cDNA was identified in an iterative
CC   search of human ESTs designed to identify basic-helix-loop-helix-PAS
CC   (bHLH-PAS) proteins that interact with either the Ah receptor (AHR)
CC   or the Ah receptor nuclear translocator (ARNT).  To obtain extended
CC   open reading frames for each EST, an anchored-PCR strategy was used
CC   to amplify additional flanking sequences from a commercial HepG2
CC   library.  MOP1 is also known as hypoxia inducible factor 1 alpha.
CC   The factor is induced by low oxygen.  It interacts with HSP90 and
CC   with ARNT.  The ARNT-dimerised factor regulates expression of
CC   erythropoietin, among other genes.  The invention also provides
CC   novel MOPs 2-9 nucleic acids (see AAX58981-88) and proteins (see
CC   AAY06289-97).  These are useful in a variety of research,
CC   diagnostic and therapeutic applications.  Several of the MOPs are
CC   alpha-class hypoxia-inducible factors.  Others are involved in
CC   circadian signal transduction.
```

```
XX
SQ   Sequence 3933 BP; 1243 A; 784 C; 750 G; 1156 T; 0 other;

  Query Match             93.6%;  Score 23.4;  DB 20;  Length 3933;
  Best Local Similarity   96.0%;  Pred. No. 0.21;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCATGTGACCATTAGGAAATGAGAG 25
          ||||||||||||| |||||||||||||
Db    676 CCATGTGACCATGAGGAAATGAGAG 700


RESULT 12
AAS85058
ID   AAS85058 standard; cDNA; 4162 BP.
XX
AC   AAS85058;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #20862.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG20871.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 20862; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4162 BP; 1286 A; 843 C; 813 G; 1220 T; 0 other;

  Query Match             93.6%;  Score 23.4;  DB 23;  Length 4162;
  Best Local Similarity   96.0%;  Pred. No. 0.21;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCATGTGACCATTAGGAAATGAGAG 25
          ||||||||||||| |||||||||||||
Db    804 CCATGTGACCATGAGGAAATGAGAG 828


RESULT 13
AAS14156
ID   AAS14156 standard; DNA; 10355 BP.
XX
AC   AAS14156;
XX
DT   18-DEC-2001  (first entry)
XX
DE   pSMART CMV-HIF DNA from a vector expressing HIF-1 alpha.
XX
KW   Differential expression; polymorphism; biological pathway; IRES; GFP; ds;
KW   internal ribosome entry site; green fluorescent protein; HIF-1 alpha;
KW   hypoxia inducible factor 1 alpha; hypoxia regulated gene; macrophage;
KW   human; CMV; cytomegalovirus.
XX
OS   Synthetic.
XX
PN   WO200162965-A2.
XX
PD   30-AUG-2001.
XX
PF   22-FEB-2001; 2001WO-GB00758.
XX
PR   22-FEB-2000; 2000GB-0004197.
PR   28-JUL-2000; 2000GB-0018679.
XX
PA   (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
PI   Kingsman AJ;
XX
DR   WPI; 2001-589807/66.
XX
PT   Screening a genetic element involved in a cellular process, comprises
PT   comparing gene expressions in a cell, and a second cell that has
PT   altered levels of genes used in the process, and detecting an element
PT   with varied expression  -
XX
PS   Example 5; Page 97-101; 103pp; English.
XX
CC   The invention relates to differential expression screening to identify a
CC   genetic element involved in a cellular process. The method involves
CC   comparing gene expressions in two cells, where one cell has altered
CC   levels of a biological molecule, and identifying the genetic element
CC   whose expression differs, by comparing expression under different
CC   environmental conditions. The method is useful for identifying mutations
CC   and polymorphisms that affect the biological response to a particular
CC   cellular process. The method also allows the molecular dissection of
CC   biological pathways by altering a particular pathway under study. By
CC   introducing a heterologous nucleic acid into a cell population, the level
CC   of a particular molecule can be influenced, allowing a pathway to be
CC   dissected into its precise molecular components. The main use of the
CC   method is to compare gene expression in cells from a diseased patient and
CC   from a normal patient. This sequence represents pSMART CMV-HIF from a
CC   vector expressing hypoxia inducible factor 1 alpha (HIF-1 alpha) which is
CC   used in methods of the invention.
XX
SQ   Sequence 10355 BP; 3060 A; 2212 C; 2373 G; 2710 T; 0 other;

  Query Match             93.6%;  Score 23.4;  DB 22;  Length 10355;
  Best Local Similarity   96.0%;  Pred. No. 0.24;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCATGTGACCATTAGGAAATGAGAG 25
```

```
          |||||||||||| ||||||||||||
Db    3305 CCATGTGACCATGAGGAAATGAGAG 3329


RESULT 14
AAK77781/c
ID   AAK77781 standard; DNA; 27884 BP.
XX
AC   AAK77781;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32593.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
```

```
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 32593; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 27884 BP; 9296 A; 5084 C; 5260 G; 8244 T; 0 other;

  Query Match             93.6%;  Score 23.4;  DB 22;  Length 27884;
  Best Local Similarity   96.0%;  Pred. No. 0.27;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||| |||||||||||||
Db   26482 CCATGTGACCATGAGGAAATGAGAG 26458


RESULT 15
AAH93977/c
ID   AAH93977 standard; cDNA; 367 BP.
XX
AC   AAH93977;
XX
DT   05-OCT-2001  (first entry)
XX
DE   Human foetal cDNA, SEQ ID NO: 506.
XX
KW   Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
KW   nootropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
KW   gene therapy; antisense therapy; cancer; immune disorder;
KW   growth disorder; osteoporosis; thrombolytic disorder;
KW   nervous system disorder; inflammation; expressed sequence tag; EST; ss.
XX
OS   Homo sapiens.
XX
PN   WO200155339-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02723.
XX
PR   25-JAN-2000; 2000US-0491404.
PR   15-SEP-2000; 2000US-0663870.
PR   06-NOV-2000; 2000US-0707351.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Yeung G,  Ford JE,  Boyle BJ,  Arterburn MC,  Drmanac RA,  Tang YT;
PI   Liu C,  Asundi V,  Zhou P,  Werhman T;
XX
DR   WPI; 2001-465571/50.
DR   P-PSDB; AAM06302.
XX
PT   Novel fetal proteins useful for the treatment and diagnosis of diseases
PT   associated with dysfunction of the protein e.g. cancers, immune
PT   disorders, growth disorders, thrombolytic disorders, nervous system
PT   disorders and inflammation -
XX
PS   Claim 1; Page 362-363; 715pp; English.
XX
CC   The invention relates to novel foetal polypeptides encoded by
CC   polynucleotides comprising one of 477 sequences fully defined in the
CC   specification. The foetal polynucleotides and polypeptides are
CC   useful in the treatment and diagnosis of diseases such as cancers,
CC   immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
CC   disorders, nervous system disorders and inflammation. The present
CC   sequence was assembled using an expressed sequence tag (EST) found
CC   to be expressed in human foetal tissue cDNA libraries as the seed.
XX
SQ   Sequence 367 BP; 94 A; 65 C; 58 G; 150 T; 0 other;

  Query Match             72.8%;  Score 18.2;  DB 22;  Length 367;
  Best Local Similarity   87.0%;  Pred. No. 44;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 ATGTGACCATTAGGAAATGAGAG 25
           ||||||| |||| ||||| ||||
Db     219 ATGTGACAATTAAGAAATAAGAG 197


Search completed: March 17, 2003, 10:50:36
Job time : 151.446 secs
```

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 10:23:12 ; Search time 883.737 Seconds
(without alignments)
458.154 Million cell updates/sec

Title: US-09-836-439-2
Perfect score: 25
Sequence: 1 ccatgtgaccattaggaaatgagag 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	23.4	93.6	337	14	BQ308862	BQ308862 MR0-BT550
	2	23.4	93.6	348	10	AW504525	AW504525 UI-HF-BN0
c	3	23.4	93.6	380	10	AW120791	AW120791 UI-M-BH2.
	4	23.4	93.6	386	10	AW124493	AW124493 UI-M-BH2.
	5	23.4	93.6	395	12	BE980521	BE980521 UI-M-BG2-
	6	23.4	93.6	501	12	BF406698	BF406698 UI-R-BJ2-
c	7	23.4	93.6	591	9	AI481068	AI481068 vf92d03.x
	8	23.4	93.6	684	14	BM834484	BM834484 K-EST0109
	9	23.4	93.6	693	10	BB199902	BB199902 BB199902
	10	23.4	93.6	711	12	BE731662	BE731662 601567048
	11	23.4	93.6	734	12	BF789080	BF789080 602104940
	12	23.4	93.6	739	13	BI413382	BI413382 602986733
	13	23.4	93.6	769	12	BG772697	BG772697 602720844
	14	23.4	93.6	791	12	BG110077	BG110077 602279683
	15	23.4	93.6	878	9	AU132294	AU132294 AU132294
	16	23.4	93.6	878	14	BQ711945	BQ711945 AGENCOURT
	17	23.4	93.6	889	14	BQ213144	BQ213144 AGENCOURT
	18	23.4	93.6	893	12	BE877244	BE877244 601485172
	19	23.4	93.6	941	14	BQ881947	BQ881947 AGENCOURT
	20	23.4	93.6	1024	14	BM807153	BM807153 AGENCOURT
	21	23.4	93.6	1657	11	AK017853	AK017853 Mus muscu
	22	22.4	89.6	714	9	AU131767	AU131767 AU131767
	23	21.8	87.2	310	10	BB204614	BB204614 BB204614
	24	21.8	87.2	903	9	AL698195	AL698195 DKFZp686O
	25	20.4	81.6	533	17	AZ786928	AZ786928 2M0032N05
	26	19.8	79.2	372	14	W86995	W86995 zh61e02.s1
	27	19.8	79.2	478	9	AA700482	AA700482 zj74f07.s
	28	19.8	79.2	1435	14	BM906863	BM906863 AGENCOURT
c	29	19.2	76.8	105	14	BQ308421	BQ308421 MR0-BT450
c	30	19.2	76.8	379	14	T31654	T31654 EST36504 Hu
c	31	19.2	76.8	415	17	AQ108979	AQ108979 CIT-HSP-2
c	32	19.2	76.8	611	17	B90780	B90780 CIT-HSP-217
	33	18.8	75.2	381	17	AZ263188	AZ263188 RPCI-23-1
c	34	18.8	75.2	622	14	BQ397883	BQ397883 NISC_mo02
	35	18.8	75.2	633	10	BB652302	BB652302 BB652302
	36	18.8	75.2	800	12	BF617665	BF617665 HVSMEc001
	37	18.6	74.4	260	17	AZ778470	AZ778470 2M0013K17
c	38	18.6	74.4	335	13	BJ198294	BJ198294 BJ198294
c	39	18.6	74.4	366	17	AZ433190	AZ433190 1M0218L22
c	40	18.6	74.4	377	13	BJ195711	BJ195711 BJ195711
c	41	18.6	74.4	425	17	AZ931901	AZ931901 474.dhz91
c	42	18.6	74.4	468	17	CNS02728	AL184121 Tetraodon
c	43	18.6	74.4	484	9	AA203974	AA203974 mu28g05.r
c	44	18.6	74.4	519	13	BM482542	BM482542 535462 MA
	45	18.6	74.4	538	17	BH026082	BH026082 RPCI-24-3

ALIGNMENTS

RESULT 1
BQ308862
LOCUS BQ308862 337 bp mRNA linear EST 16-MAY-2002
DEFINITION MR0-BT5505-040701-003-c02 BT5505 Homo sapiens cDNA, mRNA sequence.
ACCESSION BQ308862
VERSION BQ308862.1 GI:20851208
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 337)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR0&t2=MR0-BT5505-040701-003-c02&t3=2001-07-04&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 15
High quality sequence stop: 331.

FEATURES Location/Qualifiers
source 1. .337
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT5505"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 105 a 49 c 81 g 102 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 14; Length 337;
Best Local Similarity 96.0%; Pred. No. 2.7;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||||| |||||||||||
Db 259 CCATGTGACCATGAGGAAATGAGAG 283

RESULT 2
AW504525
LOCUS AW504525 348 bp mRNA linear EST 02-MAR-2000
DEFINITION UI-HF-BN0-alj-d-05-0-UI.r1 NIH_MGC_50 Homo sapiens cDNA clone IMAGE:3079736 5', mRNA sequence.
ACCESSION AW504525
VERSION AW504525.1 GI:7142192
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 348)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M.B. Soares Lab
cDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.

FEATURES Location/Qualifiers
source 1. .348
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3079736"
/clone_lib="NIH_MGC_50"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTI)"
/note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (3.5-4.4kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D. "

BASE COUNT 107 a 54 c 81 g 106 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 10; Length 348;
Best Local Similarity 96.0%; Pred. No. 2.7;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||||| |||||||||||
Db 255 CCATGTGACCATGAGGAAATGAGAG 279

RESULT 3
AW120791/c
LOCUS AW120791 380 bp mRNA linear EST 22-OCT-1999
DEFINITION UI-M-BH2.3-any-h-05-0-UI.s1 NIH_BMAP_M_S3.3 Mus musculus cDNA clone UI-M-BH2.3-any-h-05-0-UI 3', mRNA sequence.
ACCESSION AW120791
VERSION AW120791.1 GI:6096124
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 380)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized basal ganglia library cDNA Library Preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined.
Seq primer: M13 Forward
POLYA=Yes.

FEATURES Location/Qualifiers
source 1. .380
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH2.3-any-h-05-0-UI"
/clone_lib="NIH_BMAP_M_S3.3"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S3.3 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hipoccampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S3.3, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S3.3) was constructed

as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S2 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the NIH_BMAP_M_S2 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (LifeTechnologies) to generate the NIH_BMAP_M_S3.3 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S3.3
TAG_TISSUE=basal-ganglia
TAG_SEQ=TGTAC"

BASE COUNT 105 a 87 c 67 g 120 t 1 others
ORIGIN

Query Match 93.6%; Score 23.4; DB 10; Length 380;
Best Local Similarity 96.0%; Pred. No. 2.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||||| |||||||||||||
Db 72 CCATGTGACCATGAGGAAATGAGAG 48

RESULT 4
AW124493
LOCUS AW124493 386 bp mRNA linear EST 22-OCT-1999
DEFINITION UI-M-BH2.1-apo-a-05-0-UI.s1 NIH_BMAP_M_S3.1 Mus musculus cDNA clone UI-M-BH2.1-apo-a-05-0-UI 3', mRNA sequence.
ACCESSION AW124493
VERSION AW124493.1 GI:6099988
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 386)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized amygdala library cDNA Library Preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined.
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .386
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH2.1-apo-a-05-0-UI"
/clone_lib="NIH_BMAP_M_S3.1"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S3.1 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hipoccampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S3.1) was constructed as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S2 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the NIH_BMAP_M_S2 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (LifeTechnologies) to generate the NIH_BMAP_M_S3.1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S3.1
TAG_TISSUE=amygdala
TAG_SEQ=GTGAG"

BASE COUNT 106 a 72 c 91 g 117 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 10; Length 386;
Best Local Similarity 96.0%; Pred. No. 2.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||||| |||||||||||||
Db 315 CCATGTGACCATGAGGAAATGAGAG 339

RESULT 5
BE980521
LOCUS BE980521 395 bp mRNA linear EST 05-OCT-2000
DEFINITION UI-M-BG2-bck-b-03-0-UI.s1 NIH_BMAP_MSC_S1 Mus musculus cDNA clone UI-M-BG2-bck-b-03-0-UI 3', mRNA sequence.
ACCESSION BE980521
VERSION BE980521.1 GI:10648591
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 395)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the amygdala tissue cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and

```
          limited collaborative arrangements
          Seq primer: M13 Forward
          POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .395
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BG2-bck-b-03-0-UI"
                     /clone_lib="NIH_BMAP_MSC_S1"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_MSC_S1 library is a subtracted library derived
                     NIH_BMAP_MSC_N. NIH_BMAP_MSC_N was made from  mouse spinal
                     cord tissue. For a detailed description of the library
                     from which this clone was  derived, please visit our web
                     site at brainest.eng.uiowa.edu.
                     TAG_LIB=NIH_BMAP_MSC_S1
                     TAG_TISSUE=amygdala
                     TAG_SEQ=GTGAG"
BASE COUNT      108 a     73 c     93 g    121 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 12;  Length 395;
  Best Local Similarity   96.0%;  Pred. No. 2.9;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||||  ||||||||||||
Db     315 CCATGTGACCATGAGGAAATGAGAG 339


RESULT 6
BF406698
LOCUS       BF406698               501 bp    mRNA    linear   EST 28-NOV-2000
DEFINITION  UI-R-BJ2-bpo-b-02-0-UI.s1 UI-R-BJ2 Rattus norvegicus cDNA clone
            UI-R-BJ2-bpo-b-02-0-UI 3', mRNA sequence.
ACCESSION   BF406698
VERSION     BF406698.1  GI:11394673
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 25-96, >B1-F#SINE/Alu
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .501
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BJ2-bpo-b-02-0-UI"
                     /clone_lib="UI-R-BJ2"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ2
                     library is a subtracted library derived from the following
                     tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc,
                     atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15
                     dpc, AV canal at 15 dpc.  For a detailed description of
                     the library from which this clone was derived, please
                     visit our web site at ratest.eng.uiowa.edu. The
                     subtraction has been previously described in (Bonaldo,
                     Lennon and Soares, Genome Research 6:791-806, 1996)
                     TAG_SEQ=None found"
BASE COUNT      141 a    100 c    125 g    135 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 12;  Length 501;
  Best Local Similarity   96.0%;  Pred. No. 3.2;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||||  ||||||||||||
Db     473 CCATGTGACCATGAGGAAATGAGAG 497


RESULT 7
AI481068/c
LOCUS       AI481068               591 bp    mRNA    linear   EST 09-MAR-1999
DEFINITION  vf92d03.x1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:851237 3' similar to TR:O08741 O08741 HYPOXIA-INDUCIBLE
            FACTOR ONE ALPHA. ;, mRNA sequence.
ACCESSION   AI481068
VERSION     AI481068.1  GI:4374294
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:503389
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            Possible reversed clone: similarity on wrong strand
            High quality sequence stop: 472.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:851237"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                     ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTT
                     T 3'];  double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia),  digested with Not I and cloned into
                     the Not I and Eco RI sites of  the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
```

Bonaldo."
BASE COUNT 150 a 143 c 116 g 182 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 9; Length 591;
Best Local Similarity 96.0%; Pred. No. 3.4;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||| |||||||||||
Db 390 CCATGTGACCATGAGGAAATGAGAG 366

RESULT 8
BM834484
LOCUS BM834484 684 bp mRNA linear EST 06-MAR-2002
DEFINITION K-EST0109485 S11SNU1 Homo sapiens cDNA clone S11SNU1-63-B11 5', mRNA sequence.
ACCESSION BM834484
VERSION BM834484.1 GI:19190893
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 684)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 63 row: B column: 11
High quality sequence stop: 684.
FEATURES Location/Qualifiers
source 1. .684
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S11SNU1-63-B11"
/clone_lib="S11SNU1"
/sex="M"
/tissue_type="Stomach"
/cell_type="Lymphoblast-like"
/cell_line="SNU-1"
/lab_host="Top10F'"
/note="Organ: Stomach; Vector: pME18-FL3; Site_1: XhoI; Site_2: XhoI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including SfiI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized with Superscript II using SfiI oligo-dT primer. After first strand synthesis, RNA was degraded by NaOH treatment and cDNA was amplified by PCR reaction. The PCR products were digested with SfiI and cloned into DraIII- digested pME18S-FL3 vector. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."
BASE COUNT 232 a 135 c 136 g 181 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 14; Length 684;
Best Local Similarity 96.0%; Pred. No. 3.6;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||| |||||||||||
Db 75 CCATGTGACCATGAGGAAATGAGAG 99

RESULT 9
BB199902
LOCUS BB199902 693 bp mRNA linear EST 19-OCT-2001
DEFINITION BB199902 RIKEN full-length enriched, 0 day neonate thymus Mus musculus cDNA clone A430018A13 3', mRNA sequence.
ACCESSION BB199902
VERSION BB199902.2 GI:16271363
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 693)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Jun 30, 2000 this sequence version replaced gi:8864855.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
FEATURES Location/Qualifiers
source 1. .693
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A430018A13"
/clone_lib="RIKEN full-length enriched, 0 day neonate thymus"

```
                     /tissue_type="thymus"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT      213 a    100 c    141 g    239 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 10;  Length 693;
  Best Local Similarity   96.0%;  Pred. No. 3.6;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||||| |||||||||||||
Db     414 CCATGTGACCATGAGGAAATGAGAG 438


RESULT 10
BE731662
LOCUS       BE731662                711 bp    mRNA    linear   EST 15-SEP-2000
DEFINITION  601567048F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3842146 5',
            mRNA sequence.
ACCESSION   BE731662
VERSION     BE731662.1  GI:10145654
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM534  row: o  column: 11
            High quality sequence stop: 697.
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3842146"
                     /clone_lib="NIH_MGC_21"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      174 a    166 c    197 g    174 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 12;  Length 711;
  Best Local Similarity   96.0%;  Pred. No. 3.6;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||||| |||||||||||||
Db     646 CCATGTGACCATGAGGAAATGAGAG 670


RESULT 11
BF789080
LOCUS       BF789080                734 bp    mRNA    linear   EST 12-JAN-2001
DEFINITION  602104940F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4223096
            5', mRNA sequence.
ACCESSION   BF789080
VERSION     BF789080.1  GI:12094116
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 734)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9810  row: p  column: 09
            High quality sequence stop: 708.
FEATURES             Location/Qualifiers
     source          1. .734
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4223096"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      217 a    147 c    170 g    200 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 12;  Length 734;
  Best Local Similarity   96.0%;  Pred. No. 3.7;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||||| |||||||||||||
Db     306 CCATGTGACCATGAGGAAATGAGAG 330


RESULT 12
BI413382
LOCUS       BI413382                739 bp    mRNA    linear   EST 14-AUG-2001
DEFINITION  602986733F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5142800 5',
            mRNA sequence.
ACCESSION   BI413382
VERSION     BI413382.1  GI:15174305
KEYWORDS    EST.
SOURCE      house mouse.
```

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 739)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by:Incyte Genomics, Inc.
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11352  row: a  column: 09
            High quality sequence start: 125
            High quality sequence stop: 726.
FEATURES              Location/Qualifiers
     source           1. .739
                      /organism="Mus musculus"
                      /strain="CZECH II"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:5142800"
                      /clone_lib="NCI_CGAP_Lu33"
                      /tissue_type="pooled lung tumors"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                      modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
                      strand cDNA was prepared from mRNA obtained from pooled
                      lung tumors with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCTGTTTTTTTTTTTTTTTT 3']. 
                      Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not
                      I and Eco RI sites of  the modified pT7T3 vector. Library
                      went through one round of normalization, and was
                      constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      180 a    176 c    210 g    171 t      2 others
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 13;  Length 739;
  Best Local Similarity   96.0%;  Pred. No. 3.7;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||||| ||||||||||||
Db     681 CCATGTGACCATGAGGAAATGAGAG 705


RESULT 13
BG772697
LOCUS       BG772697                 769 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602720844F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4837691 5',
            mRNA sequence.
ACCESSION   BG772697
VERSION     BG772697.1  GI:14083350
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 769)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10770  row: h  column: 12
            High quality sequence stop: 723.
FEATURES              Location/Qualifiers
     source           1. .769
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4837691"
                      /clone_lib="NIH_MGC_97"
                      /lab_host="DH10B"
                      /note="Organ: testis; Vector: pBluescriptR (modified
                      pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                      ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
                      size-selected for  average insert size 2.2 kb and
                      normalized to ROT 5. This is a primary library enriched
                      for full-length clones and constructed using the
                      Cap-trapper method (Carninci, in preparation).  Library
                      constructed by M. Brownstein (NIMH/NHGRI, National
                      Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      243 a    143 c    170 g    213 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 12;  Length 769;
  Best Local Similarity   96.0%;  Pred. No. 3.8;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCATGTGACCATTAGGAAATGAGAG 25
           ||||||||||||| ||||||||||||
Db     434 CCATGTGACCATGAGGAAATGAGAG 458


RESULT 14
BG110077
LOCUS       BG110077                 791 bp    mRNA    linear   EST 30-JAN-2001
DEFINITION  602279683F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4367229 5',
            mRNA sequence.
ACCESSION   BG110077
VERSION     BG110077.1  GI:12603583
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10019  row: e  column: 22
            High quality sequence stop: 715.
FEATURES              Location/Qualifiers
     source           1. .791
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4367229"
                      /clone_lib="NIH_MGC_86"
                      /tissue_type="osteosarcoma, cell line"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: bone; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                      Average insert size 1.533 kb. Library enriched for
                      full-length clones and constructed by Life Technologies.
```

Note: this is a NIH_MGC Library."
BASE COUNT 250 a 145 c 173 g 223 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 12; Length 791;
Best Local Similarity 96.0%; Pred. No. 3.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||||| |||||||||||||
Db 236 CCATGTGACCATGAGGAAATGAGAG 260

RESULT 15
AU132294
LOCUS AU132294 878 bp mRNA linear EST 01-AUG-2002
DEFINITION AU132294 NT2RP3 Homo sapiens cDNA clone NT2RP3004165 5', mRNA sequence.
ACCESSION AU132294
VERSION AU132294.1 GI:10992648
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 878)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.
FEATURES Location/Qualifiers
source 1. .878
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP3004165"
/clone_lib="NT2RP3"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 274 a 163 c 188 g 248 t 5 others
ORIGIN

Query Match 93.6%; Score 23.4; DB 9; Length 878;
Best Local Similarity 96.0%; Pred. No. 4;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATTAGGAAATGAGAG 25
||||||||||||| |||||||||||||
Db 375 CCATGTGACCATGAGGAAATGAGAG 399

Search completed: March 17, 2003, 13:09:12
Job time : 888.737 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:43:42 ; Search time 220.903 Seconds
(without alignments)
3161.870 Million cell updates/sec

Title: US-09-836-439-3
Perfect score: 24
Sequence: 1 gctttctttgccaagagcgccgca 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21.4	89.2	175	5	AF249160	AF249160 Philautus
2	21.4	89.2	1056	5	SPIY18677	Y18677 Sardina pil
3	21.4	89.2	2577	5	CLRHI1	AF149230 Columba l
4	21.4	89.2	3016	6	I18746	I18746 Sequence 1
5	21.4	89.2	5385	9	AB065668	AB065668 Homo sapi
6	21.4	89.2	6953	6	I18747	I18747 Sequence 2
7	21.4	89.2	6953	9	HSU49742	U49742 Human rhodo
8	21.4	89.2	163297	2	AC023162	AC023162 Homo sapi
9	21.4	89.2	164396	2	HSAC000380	AC000380 Homo sapi
c 10	21.4	89.2	168551	9	AC080007	AC080007 Homo sapi
11	20.4	85.0	867	5	U97272	U97272 Abyssocottu
12	19.8	82.5	175	5	AF249130	AF249130 Hyperoliu
13	19.8	82.5	175	5	AF249133	AF249133 Mantella
14	19.8	82.5	175	5	AF249134	AF249134 Mantidact
15	19.8	82.5	175	5	AF249136	AF249136 Boophis x
16	19.8	82.5	175	5	AF249137	AF249137 Boophis t
17	19.8	82.5	175	5	AF249138	AF249138 Laliostom
18	19.8	82.5	175	5	AF249139	AF249139 Fejervary
19	19.8	82.5	175	5	AF249140	AF249140 Fejervary
20	19.8	82.5	175	5	AF249141	AF249141 Hoplobatr
21	19.8	82.5	175	5	AF249142	AF249142 Sphaeroth
22	19.8	82.5	175	5	AF249143	AF249143 Euphlycti
23	19.8	82.5	175	5	AF249144	AF249144 Nannophry
24	19.8	82.5	175	5	AF249146	AF249146 Nyctibatr
25	19.8	82.5	175	5	AF249147	AF249147 Limnonect
26	19.8	82.5	175	5	AF249148	AF249148 Limnonect
27	19.8	82.5	175	5	AF249149	AF249149 Rana curt
28	19.8	82.5	175	5	AF249150	AF249150 Rana temp
29	19.8	82.5	175	5	AF249151	AF249151 Rana temp
30	19.8	82.5	175	5	AF249152	AF249152 Micrixalu
31	19.8	82.5	175	5	AF249154	AF249154 Indirana
32	19.8	82.5	175	5	AF249156	AF249156 Polypedat
33	19.8	82.5	175	5	AF249157	AF249157 Rhacophor
34	19.8	82.5	175	5	AF249158	AF249158 Philautus
35	19.8	82.5	726	5	AF369050	AF369050 Polyodon
36	19.8	82.5	858	5	AF137208	AF137208 Amia calv
37	19.8	82.5	864	5	AF137206	AF137206 Acipenser
38	19.8	82.5	867	5	U97265	U97265 Cottocomeph
39	19.8	82.5	867	5	U97266	U97266 Cottocomeph
40	19.8	82.5	867	5	U97274	U97274 Comephorus
41	19.8	82.5	1047	4	AF055319	AF055319 Trichechu
42	19.8	82.5	1053	5	AF021242	AF021242 Melopsitt
43	19.8	82.5	1150	4	AF008947	AF008947 Sus scrof
44	19.8	82.5	1300	10	AF309568	AF309568 Spalax eh
45	19.8	82.5	1389	5	PMIY18679	Y18679 Pomatoschis

ALIGNMENTS

RESULT 1
AF249160
LOCUS AF249160 175 bp DNA linear VRT 17-JAN-2001
DEFINITION Philautus charius rhodopsin gene, exon 4 and partial cds.
ACCESSION AF249160
VERSION AF249160.1 GI:12247249
KEYWORDS .
SOURCE Philautus charius.
ORGANISM Philautus charius
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Rhacophoridae;
Philautus.
REFERENCE 1 (bases 1 to 175)
AUTHORS Bossuyt,F. and Milinkovitch,M.C.
TITLE Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs

```
            Reveal Co-variation between Larval and Adult Traits
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
            Molecular Biology and Medicine, rue Jeener and Brachet 12,
            Gosselies B-6041, Belgium
FEATURES             Location/Qualifiers
     source          1. .175
                     /organism="Philautus charius"
                     /db_xref="taxon:129025"
     mRNA            <1. .>175
                     /product="rhodopsin"
     CDS             <1. .>175
                     /codon_start=2
                     /product="rhodopsin"
                     /protein_id="AAG49803.1"
                     /db_xref="GI:12247250"
                     /translation="AEKEVTRMVVIMVVFFLICWVPYAYVAFYIFTHQGSEFGPIFMT
                     VPAFFAKSSAIYNP"
     exon            <1. .>175
                     /number=4
BASE COUNT       32 a     55 c     36 g     52 t
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 5;  Length 175;
  Best Local Similarity   95.7%;  Pred. No. 5.5;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCCAAGAGCGCCGC 23
          |||||||||||||||||| ||||
Db    140 GCTTTCTTTGCCAAGAGCTCCGC 162


RESULT 2
SPIY18677
LOCUS       SPIY18677               1056 bp    mRNA    linear   VRT 26-MAY-1999
DEFINITION  Sardina pilchardus mRNA for opsin.
ACCESSION   Y18677
VERSION     Y18677.1  GI:4210840
KEYWORDS    opsin.
SOURCE      Sardina pilchardus.
  ORGANISM  Sardina pilchardus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Clupeomorpha; Clupeidae;
            Sardina.
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   Archer,S.N. and Hirano,J.
  TITLE     Comparative analysis of opsins in Mediterranian coastal fish
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1056)
  AUTHORS   Archer,S.N.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-1999) S.N. Archer, International Marine Centre,
            Localita sa Mardini, 09072 Torregrande, Oristano, ITALY
FEATURES             Location/Qualifiers
     source          1. .1056
                     /organism="Sardina pilchardus"
                     /db_xref="taxon:27697"
                     /tissue_type="retina"
                     /dev_stage="adult"
     CDS             1. .1056
                     /function="phototransduction"
                     /codon_start=1
                     /product="opsin"
                     /protein_id="CAA77259.1"
                     /db_xref="GI:4210841"
                     /translation="MNGTEGPFFYIPMSNATGLVRSPYDYPQYYLVPPWGYACLAAYM
                     FLLILTGFPVNFLTLYVTIEHKKLRSPLNYILLNLAVADLFMVIGGFTTTMWTSLNGY
                     FVFGRMGCNIEGFFATLGGEIALWSLVVLSMERWIVVCKPISNFRFGENHAVMGVAFS
                     WFMAAACAVPPLVGWSRYIPEGMQCSCGIDYYTRAEGFNNESFVIYMFVVHFTCPLTI
                     ITFCYGRLVCTVKEAAAQQQESETTQRAEREVTRMVIIMFVAFLACWVPYASVAWYIF
                     THQGSEFGPVFMTIPAFFAKSSAVYNPVIYICLNKQFRHCMITTLCCGKNPFEEEEGS
                     TTASKTEASSVCSVSPA"
     misc_feature    1. .17
                     /note="degenerate primer"
     misc_feature    1046. .1056
                     /note="degenerate primer"
BASE COUNT      186 a    368 c    269 g    233 t
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 5;  Length 1056;
  Best Local Similarity   95.7%;  Pred. No. 6.6;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCCAAGAGCGCCGC 23
          |||||||||||||||||| ||||
Db    874 GCTTTCTTTGCCAAGAGCTCCGC 896


RESULT 3
CLRHI1
LOCUS       CLRHI1                  2577 bp    DNA     linear   VRT 17-DEC-1999
DEFINITION  Columba livia RH1 opsin (rh1) gene, exons 1 through 4.
ACCESSION   AF149230
VERSION     AF149230.1  GI:4887218
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      Columba livia.
  ORGANISM  Columba livia
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Columbiformes; Columbidae; Columba.
REFERENCE   1  (bases 1 to 2577)
  AUTHORS   Kawamura,S., Blow,N.S. and Yokoyama,S.
  TITLE     Genetic analyses of visual pigments of the pigeon (Columba livia)
  JOURNAL   Genetics 153 (4), 1839-1850 (1999)
  MEDLINE   20050679
   PUBMED   10581289
REFERENCE   2  (bases 1 to 2577)
  AUTHORS   Yokoyama,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-1999) Department of Biology, Syracuse University,
            130 College Place, Syracuse, NY 13244, USA
FEATURES             Location/Qualifiers
     source          1. .2577
                     /organism="Columba livia"
                     /db_xref="taxon:8932"
     exon            <76. .436
                     /gene="rh1"
                     /number=1
     exon            1293. .1461
                     /gene="rh1"
                     /number=2
     exon            1561. .1726
                     /gene="rh1"
                     /number=3
     exon            1968. .2207
                     /gene="rh1"
                     /number=4
BASE COUNT      442 a    778 c    815 g    542 t
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 5;  Length 2577;
  Best Local Similarity   95.7%;  Pred. No. 7.1;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCCAAGAGCGCCGC 23
          |||||||||||||||||| ||||
Db   2145 GCTTTCTTTGCCAAGAGCTCCGC 2167


RESULT 4
I18746
LOCUS       I18746                  3016 bp    DNA     linear   PAT 07-OCT-1996
DEFINITION  Sequence 1 from patent US 5498521.
```

```
ACCESSION   I18746
VERSION     I18746.1  GI:1599101
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3016)
  AUTHORS   Dryja,T.P. and Berson,E.L.
  TITLE     Diagnosis of hereditary retinal degenerative diseases
  JOURNAL   Patent: US 5498521-A 1 12-MAR-1996;
FEATURES             Location/Qualifiers
     source          1. .3016
                     /organism="unknown"
BASE COUNT      689 a    863 c    753 g    711 t
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 6;  Length 3016;
  Best Local Similarity   95.7%;  Pred. No. 7.2;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCCAAGAGCGCCGC 23
          || ||||||||||||||||||||
Db   1168 GCGTTCTTTGCCAAGAGCGCCGC 1190


RESULT 5
AB065668
LOCUS       AB065668                5385 bp    DNA     linear   PRI 23-JUL-2002
DEFINITION  Homo sapiens gene for seven transmembrane helix receptor, complete
            cds, isolate:CBRC7TM_231.
ACCESSION   AB065668
VERSION     AB065668.1  GI:21928610
KEYWORDS    .
SOURCE      Homo sapiens (isolate:CBRC7TM_231) DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Suwa,M., Sato,T., Okouchi,I., Arita,M., Futami,K., Matsumoto,S.,
            Tsutsumi,S., Aburatani,H., Asai,K. and Akiyama,Y.
  TITLE     Genome-wide discovery and analysis of human seven transmembrane
            helix receptor genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5385)
  AUTHORS   Suwa,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2001) Makiko Suwa, Computational Biology Research
            Center (CBRC), National Institute of Advanced Industrial Science
            and Technology (AIST); 2-41-6 Aomi Koto-ku, Tokyo 135-0064, Japan
            (E-mail:m-suwa@aist.go.jp, URL:http://www.cbrc.jp/,
            Tel:81-3-3599-8080, Fax:81-3-3599-8081)
COMMENT     This sequence is a seven transmembrane helix receptor candidate
            predicted from the whole human genome sequences using our automated
            system that contains programs of gene
            finding(GeneDecoder), sequence search, motif-domain assignment and
            transmembrane helix prediction.
            And the sequence is submitted by the collaborative project between
            [Computational Biology Research Center (CBRC), National Institute
            of Advanced Industrial Science and Technology (AIST)] and [Genome
            Science Division, Research Center for Advanced Science and
            Technology (RCAST), University of Tokyo].
FEATURES             Location/Qualifiers
     source          1. .5385
                     /organism="Homo sapiens"
                     /isolate="CBRC7TM_231"
                     /db_xref="taxon:9606"
                     /chromosome="3"
     CDS             join(201. .561,2343. .2511,3718. .3883,4000. .4239,5075. .5185)
                     /codon_start=1
                     /evidence=not_experimental
                     /product="seven transmembrane helix receptor"
                     /protein_id="BAC05894.1"
                     /db_xref="GI:21928611"
                     /translation="MNGTEGPNFYVPFSNATGVVRSPFEYPQYYLAEPWQFSMLAAYM
                     FLLIVLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFMVLGGFTSTLYTSLHGY
                     FVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFT
                     WVMALACAAPPLAGWSRYIPEGLQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPMII
                     IFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWVPYASVAFYIF
                     THQGSNFGPIFMTIPAFFAKSAAIYNPVIYIMMNKQFRNCMLTTICCGKNPLGDDEAS
                     ATVSKTETSQVAPA"
BASE COUNT      1132 a    1618 c    1407 g    1228 t
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 9;  Length 5385;
  Best Local Similarity   95.7%;  Pred. No. 7.7;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCCAAGAGCGCCGC 23
          || ||||||||||||||||||||
Db   4177 GCGTTCTTTGCCAAGAGCGCCGC 4199


RESULT 6
I18747
LOCUS       I18747                  6953 bp    DNA     linear   PAT 07-OCT-1996
DEFINITION  Sequence 2 from patent US 5498521.
ACCESSION   I18747
VERSION     I18747.1  GI:1599102
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6953)
  AUTHORS   Dryja,T.P. and Berson,E.L.
  TITLE     Diagnosis of hereditary retinal degenerative diseases
  JOURNAL   Patent: US 5498521-A 2 12-MAR-1996;
FEATURES             Location/Qualifiers
     source          1. .6953
                     /organism="unknown"
BASE COUNT      1523 a    2022 c    1797 g    1611 t
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 6;  Length 6953;
  Best Local Similarity   95.7%;  Pred. No. 7.8;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCCAAGAGCGCCGC 23
          || ||||||||||||||||||||
Db   4272 GCGTTCTTTGCCAAGAGCGCCGC 4294


RESULT 7
HSU49742
LOCUS       HSU49742                6953 bp    DNA     linear   PRI 14-AUG-1998
DEFINITION  Human rhodopsin gene, complete cds.
ACCESSION   U49742 K02281
VERSION     U49742.1  GI:1236136
KEYWORDS    opsin; rhodopsin.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6953)
  AUTHORS   Nathans,J. and Hogness,D.S.
  TITLE     Isolation and nucleotide sequence of the gene encoding human
            rhodopsin
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 81 (15), 4851-4855 (1984)
  MEDLINE   84272729
   PUBMED   6589631
REFERENCE   2  (bases 1 to 6953)
  AUTHORS   Nathans,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1996) Jeremy Nathans, Molecular Biology and
            Genetics, Johns Hopkins Medical School, 725 N. Wolfe Street,
            Baltimore, MD 21205, USA
COMMENT     On Sep 3, 1996 this sequence version replaced gi:189393.
```

```
FEATURES             Location/Qualifiers
     source          1. .6953
                     /organism="Homo sapiens"
                     /isolate="individual J.N."
                     /db_xref="taxon:9606"
                     /clone="roJHN"
     CAAT_signal     122. .127
                     /note="putative CAAT box"
     TATA_signal     171. .177
                     /note="putative TATA box"
     prim_transcript 200. .>5278
                     /note="rhodopsin mRNA (alt); the 5' untranslated region is
                     similar to the bovine sequence"
     prim_transcript 202. .>5278
                     /note="rhodopsin mRNA (alt); the 5' untranslated region is
                     similar to the bovine sequence"
     CDS             join(295. .655,2439. .2607,3813. .3978,4095. .4334,5168. .5278)
                     /note="the 5 exons of the human rhodopsin gene are similar
                     to the 5 exons of the Bos taurus opsin gene"
                     /codon_start=1
                     /product="rhodopsin"
                     /protein_id="AAC31763.1"
                     /db_xref="GI:1236137"
                     /translation="MNGTEGPNFYVPFSNATGVVRSPFEYPQYYLAEPWQFSMLAAYM
                     FLLIVLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFMVLGGFTSTLYTSLHGY
                     FVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFT
                     WVMALACAAPPLAGWSRYIPEGLQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPMII
                     IFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWVPYASVAFYIF
                     THQGSNFGPIFMTIPAFFAKSAAIYNPVIYIMMNKQFRNCMLTTICCGKNPLGDDEAS
                     ATVSKTETSQVAPA"
     exon            <295. .655
                     /number=1
     intron          656. .2438
                     /note="intron A; the four introns present in both the
                     human and the bovine genes occur at precisely analogous
                     positions and are of comparable lengths"
     exon            2439. .2607
                     /number=2
     intron          2608. .3812
                     /note="intron B; the four introns present in both the
                     human and the bovine genes occur at precisely analogous
                     positions and are of comparable lengths"
     exon            3813. .3978
                     /number=3
     intron          3979. .4094
                     /note="intron C; the four introns present in both the
                     human and the bovine genes occur at precisely analogous
                     positions and are of comparable lengths"
     exon            4095. .4334
                     /number=4
     intron          4335. .5167
                     /note="intron D; the four introns present in both the
                     human and the bovine genes occur at precisely analogous
                     positions and are of comparable lengths"
     exon            5168. .>5278
                     /number=5
     polyA_signal    5642. .5647
                     /note="potential polyA signal"
     polyA_signal    6698. .6903
                     /note="potential polyA signal"
BASE COUNT     1524 a   2022 c   1796 g   1611 t
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 9;  Length 6953;
  Best Local Similarity   95.7%;  Pred. No. 7.8;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || |||||||||||||||||||||
Db    4272 GCGTTCTTTGCCAAGAGCGCCGC 4294


RESULT 8
```

```
AC023162
LOCUS       AC023162              163297 bp    DNA     linear   HTG 24-JUL-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-263I1, WORKING DRAFT SEQUENCE,
            4 unordered pieces.
ACCESSION   AC023162
VERSION     AC023162.30  GI:21908269
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163297)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163297)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 163297)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 19, 2002 this sequence version replaced gi:20127313.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Drafting Center Code: BCM
                Contact: hgsc-help@bcm.tmc.edu
```

```
            ---------- Project Information
                Center project name: HAEW
                Center clone name: RP11-263I1
            ---------- Summary Statistics
                Sequencing vector: Plasmid;
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 12% of reads
                Chemistry: Dye-terminator Big Dye: 87% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 161558 bases at least Q40
                Consensus quality: 162613 bases at least Q30
                Consensus quality: 163127 bases at least Q20
                Estimated insert size: 312099; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2633: contig of 2633 bp in length
            *     2634     2733: gap of unknown length
            *     2734    22721: contig of 19988 bp in length
            *    22722    22821: gap of unknown length
            *    22822    57017: contig of 34196 bp in length
            *    57018    57117: gap of unknown length
            *    57118   163297: contig of 106180 bp in length.
FEATURES                Location/Qualifiers
     source             1. .163297
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="3"
                        /clone="RP11-263I1"
BASE COUNT     38049 a  44043 c  43134 g  37761 t    310 others
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 2;  Length 163297;
  Best Local Similarity   95.7%;  Pred. No. 11;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || ||||||||||||||||||||
Db   89177 GCGTTCTTTGCCAAGAGCGCCGC 89199


RESULT 9
HSAC000380
LOCUS       HSAC000380            164396 bp    DNA     linear   HTG 13-JUN-2002
DEFINITION  Homo sapiens chromosome 3 clone pDJ70i11, *** SEQUENCING IN
            PROGRESS ***, 2 unordered pieces.
ACCESSION   AC000380
VERSION     AC000380.2  GI:21406173
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164396)
  AUTHORS   Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Davie,J.,
            Davies,C.J., Davis,C., English,C., Fondon,T., Franklin,T.L.,
            Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J.,
            Hinson,S., Megarity,C., Narayanaswamy,U., Newton,J., O'Brien,K.,
            Oliver,T., Patel,P., Probst,S., Rayner,S., Schageman,J.,
            Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T.,
            Wilson,R. and Burbee,D.
  TITLE     Title
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164396)
  AUTHORS   Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Davie,J.,
            Davies,C.J., Davis,C., English,C., Fondon,T., Franklin,T.L.,
            Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J.,
            Hinson,S., Megarity,C., Narayanaswamy,U., Newton,J., O'Brien,K.,
            Oliver,T., Patel,P., Probst,S., Rayner,S., Schageman,J.,
            Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T.,
            Wilson,R. and Burbee,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-1997) Genome Science and Technology Center,
            University of Texas Southwestern Medical Center at Dallas, 5323
            Harry Hines Blvd, Dallas, TX 75235-8591, USA
COMMENT     On Jun 13, 2002 this sequence version replaced gi:1907283.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7834: contig of 7834 bp in length
            *     7835  7934: gap of      100 bp
            *     7935   164396: contig of 156462 bp in length.
FEATURES                Location/Qualifiers
     source             1. .164396
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="3"
                        /clone="pDJ70i11"
BASE COUNT     38527 a  43485 c  42989 g  39285 t    110 others
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 2;  Length 164396;
  Best Local Similarity   95.7%;  Pred. No. 11;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || ||||||||||||||||||||
Db   85746 GCGTTCTTTGCCAAGAGCGCCGC 85768


RESULT 10
AC080007/c
LOCUS       AC080007              168551 bp    DNA     linear   PRI 07-MAR-2002
DEFINITION  Homo sapiens 3q BAC RP11-529F4 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC080007
VERSION     AC080007.26  GI:17154995
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168551)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
```

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 168551)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 168551)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 4 (bases 1 to 168551)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Nov 30, 2001 this sequence version replaced gi:16572867.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT.

FEATURES	Location/Qualifiers
source	1. .168551
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="3"
	/clone="RP11-529F4"
repeat_region	complement(1767. .2096)
	/rpt_family="L2"
repeat_region	complement(2297. .2434)
	/rpt_family="FLAM_C"
repeat_region	2436. .2537
	/rpt_family="MIR"
repeat_region	3181. .3356
	/rpt_family="MIR"
repeat_region	complement(4272. .4567)
	/rpt_family="AluJo"
repeat_region	complement(4732. .4796)
	/rpt_family="MIR"
repeat_region	5772. .5848
	/rpt_family="CT-rich"
repeat_region	5868. .6028
	/rpt_family="(TCTCTG)n"
repeat_region	7237. .7381
	/rpt_family="MIR"
repeat_region	complement(7437. .7576)
	/rpt_family="MIR"
repeat_region	complement(8862. .8938)
	/rpt_family="L2"
STS	11155. .11255
	/standard_name="50875"
STS	11160. .11259
	/standard_name="503"
repeat_region	complement(11913. .12058)
	/rpt_family="MIR"
repeat_region	13734. .13986
	/rpt_family="L2"
repeat_region	complement(14132. .14611)
	/rpt_family="MLT1G1"
repeat_region	14626. .14774
	/rpt_family="L2"
repeat_region	15040. .15150
	/rpt_family="L2"
repeat_region	15237. .15293
	/rpt_family="L2"
repeat_region	15749. .15821
	/rpt_family="MIR"
repeat_region	16119. .16298
	/rpt_family="MIR"
repeat_region	complement(16384. .16612)
	/rpt_family="AluJb"
repeat_region	16740. .16824
	/rpt_family="L2"
repeat_region	complement(16825. .17126)
	/rpt_family="AluSx"
repeat_region	17127. .17206
	/rpt_family="L2"
repeat_region	complement(18965. .19068)
	/rpt_family="L2"
repeat_region	complement(19582. .19880)
	/rpt_family="AluSg"
repeat_region	19996. .20455
	/rpt_family="L1MC5"
repeat_region	complement(20456. .20629)
	/rpt_family="AluJb"

```
     repeat_region   20630. .20687
                     /rpt_family="(TTA)n"
     repeat_region   complement(20688. .20803)
                     /rpt_family="AluJb"
     repeat_region   20804. .20825
                     /rpt_family="L1MC5"
     repeat_region   20829. .21366
                     /rpt_family="L1MB8"
     repeat_region   21526. .21735
                     /rpt_family="AluJb"
     repeat_region   22545. .22816
                     /rpt_family="L2"
     repeat_region   22835. .23066
                     /rpt_family="L2"
     repeat_region   23334. .23387
                     /rpt_family="(TGGG)n"
     repeat_region   23766. .23854
                     /rpt_family="MIR"
     repeat_region   24187. .24365
                     /rpt_family="L1MC3"
     repeat_region   24459. .24748
                     /rpt_family="AluSq"
     repeat_region   24751. .24785
                     /rpt_family="(TAAAA)n"
     repeat_region   24814. .25130

  Query Match             89.2%;  Score 21.4;  DB 9;  Length 168551;
  Best Local Similarity   95.7%;  Pred. No. 11;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || ||||||||||||||||||||
Db   34088 GCGTTCTTTGCCAAGAGCGCCGC 34066


RESULT 11
U97272
LOCUS       U97272                   867 bp    DNA     linear   VRT 12-AUG-1999
DEFINITION  Abyssocottus korotneffi rod opsin gene, partial cds.
ACCESSION   U97272
VERSION     U97272.1  GI:2226036
KEYWORDS    .
SOURCE      Abyssocottus korotneffi.
  ORGANISM  Abyssocottus korotneffi
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Scorpaeniformes;
            Cottoidei; Abyssocottidae; Abyssocottus.
REFERENCE   1  (bases 1 to 867)
  AUTHORS   Hunt,D.M., Fitzgibbon,J., Slobodyanyuk,S.J., Bowmaker,J.K. and
            Dulai,K.S.
  TITLE     Molecular evolution of the cottoid fish endemic to Lake Baikal
            deduced from nuclear DNA evidence
  JOURNAL   Mol. Phylogenet. Evol. 8 (3), 415-422 (1997)
  MEDLINE   98086781
   PUBMED   9417898
REFERENCE   2  (bases 1 to 867)
  AUTHORS   Hunt,D.M., Fitzgibbon,J., Slobodyanyuk,S.J., Bowmaker,J.K. and
            Dulai,K.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-1997) Molecular Genetics, Ophthalmology, Bath
            Street, London EC1V 9EL, UK
FEATURES             Location/Qualifiers
     source          1. .867
                     /organism="Abyssocottus korotneffi"
                     /db_xref="taxon:61637"
     CDS             <1. .>867
                     /codon_start=1
                     /product="rod opsin"
                     /protein_id="AAB61726.1"
                     /db_xref="GI:2226037"
                     /translation="YLVNPAAYAALGAYMFLLILIGFPINFLTLYVTLEHKKLRTPLN
                     YILLNLAVANLFMVLGGFTTTMYTSMHGYFVLGRLGCNLEAFFATLGGEIALWSLVVL
                     AIERWIVVCKPISNFRFTEDHAIMGLAFTWVMALACAVPPLVGWSRYIPEGMQCSCGV
                     DYYTRAEGFNNESFVIYMFIVHFLIPLSVIFFCYGRLLCAVKEAPAAQQESETTQRAE
                     KEVSRMVVIMVIGFLVCWLPYASVAWWIFCNQGSDFGPIFMTLPSFFAKSAAIYNPMI
                     YICMNKQFRHCMI"
BASE COUNT      152 a    293 c    223 g    199 t
ORIGIN

  Query Match             85.0%;  Score 20.4;  DB 5;  Length 867;
  Best Local Similarity   95.5%;  Pred. No. 21;
  Matches   21;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 CTTTCTTTGCCAAGAGCGCCGC 23
           | ||||||||||||||||||||
Db     788 CCTTCTTTGCCAAGAGCGCCGC 809


RESULT 12
AF249130
LOCUS       AF249130                 175 bp    DNA     linear   VRT 17-JAN-2001
DEFINITION  Hyperolius sp. rhodopsin gene, exon 4 and partial cds.
ACCESSION   AF249130
VERSION     AF249130.1  GI:12247189
KEYWORDS    .
SOURCE      Hyperolius sp.
  ORGANISM  Hyperolius sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Hyperoliidae;
            Hyperolius.
REFERENCE   1  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
            Reveal Co-variation between Larval and Adult Traits
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
            Molecular Biology and Medicine, rue Jeener and Brachet 12,
            Gosselies B-6041, Belgium
FEATURES             Location/Qualifiers
     source          1. .175
                     /organism="Hyperolius sp."
                     /db_xref="taxon:143655"
     mRNA            <1. .>175
     CDS             /product="rhodopsin"
                     <1. .>175
                     /codon_start=2
                     /product="rhodopsin"
                     /protein_id="AAG49773.1"
                     /db_xref="GI:12247190"
                     /translation="AEKEVTRMCIIMVVFFCISWFPYAFVAFWIFTHQGSEFGPIFMT
                     VPAFFAKSSSVYNP"
     exon            <1. .>175
                     /number=4
BASE COUNT       30 a     48 c     37 g     59 t      1 others
ORIGIN

  Query Match             82.5%;  Score 19.8;  DB 5;  Length 175;
  Best Local Similarity   91.3%;  Pred. No. 37;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           |||||||||||||||||  || |
Db     140 GCTTTCTTTGCCAAGAGCTCCTC 162


RESULT 13
AF249133
LOCUS       AF249133                 175 bp    DNA     linear   VRT 17-JAN-2001
DEFINITION  Mantella madagascariensis rhodopsin gene, exon 4 and partial cds.
ACCESSION   AF249133
VERSION     AF249133.1  GI:12247195
KEYWORDS    .
```

```
SOURCE      Mantella madagascariensis.
  ORGANISM  Mantella madagascariensis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Rhacophoridae;
            Mantella.
REFERENCE   1  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
            Reveal Co-variation between Larval and Adult Traits
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
            Molecular Biology and Medicine, rue Jeener and Brachet 12,
            Gosselies B-6041, Belgium
FEATURES             Location/Qualifiers
     source          1. .175
                     /organism="Mantella madagascariensis"
                     /db_xref="taxon:111290"
     mRNA            <1. .>175
                     /product="rhodopsin"
     CDS             <1. .>175
                     /codon_start=2
                     /product="rhodopsin"
                     /protein_id="AAG49776.1"
                     /db_xref="GI:12247196"
                     /translation="AEKEVTRMVVIMVVFFLICWVPYAYVAFYIFTHQGSEFGPIFMT
                     VPAFFAKSSAIYNP"
     exon            <1. .>175
                     /number=4
BASE COUNT       33 a     53 c     36 g     52 t      1 others
ORIGIN

  Query Match             82.5%;  Score 19.8;  DB 5;  Length 175;
  Best Local Similarity   91.3%;  Pred. No. 37;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           |||||||||||||||||| | ||
Db     140 GCTTTCTTTGCCAAGAGCTCTGC 162


RESULT 14
AF249134
LOCUS       AF249134                175 bp    DNA     linear   VRT 17-JAN-2001
DEFINITION  Mantidactylus cf. ulcerosus rhodopsin gene, exon 4 and partial cds.
ACCESSION   AF249134
VERSION     AF249134.1  GI:12247197
KEYWORDS    .
SOURCE      Mantidactylus cf. ulcerosus.
  ORGANISM  Mantidactylus cf. ulcerosus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Rhacophoridae;
            Mantidactylus.
REFERENCE   1  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
            Reveal Co-variation between Larval and Adult Traits
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
            Molecular Biology and Medicine, rue Jeener and Brachet 12,
            Gosselies B-6041, Belgium
FEATURES             Location/Qualifiers
     source          1. .175
                     /organism="Mantidactylus cf. ulcerosus"
                     /db_xref="taxon:129014"
     mRNA            <1. .>175
                     /product="rhodopsin"
     CDS             <1. .>175
                     /codon_start=2
                     /product="rhodopsin"
                     /protein_id="AAG49777.1"
                     /db_xref="GI:12247198"
                     /translation="AEKEVTRMVVIMVVFFLICWVPYAYVAFYIFTHQGSEFGPIFMT
                     VPAFFAKSSAIYNP"
     exon            <1. .>175
                     /number=4
BASE COUNT       36 a     50 c     36 g     53 t
ORIGIN

  Query Match             82.5%;  Score 19.8;  DB 5;  Length 175;
  Best Local Similarity   91.3%;  Pred. No. 37;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           |||||||||||||||||| | ||
Db     140 GCTTTCTTTGCCAAGAGCTCTGC 162


RESULT 15
AF249136
LOCUS       AF249136                175 bp    DNA     linear   VRT 17-JAN-2001
DEFINITION  Boophis xerophilus rhodopsin gene, exon 4 and partial cds.
ACCESSION   AF249136
VERSION     AF249136.1  GI:12247201
KEYWORDS    .
SOURCE      Boophis xerophilus.
  ORGANISM  Boophis xerophilus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Rhacophoridae;
            Boophis.
REFERENCE   1  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
            Reveal Co-variation between Larval and Adult Traits
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
            Molecular Biology and Medicine, rue Jeener and Brachet 12,
            Gosselies B-6041, Belgium
FEATURES             Location/Qualifiers
     source          1. .175
                     /organism="Boophis xerophilus"
                     /db_xref="taxon:128996"
     mRNA            <1. .>175
                     /product="rhodopsin"
     CDS             <1. .>175
                     /codon_start=2
                     /product="rhodopsin"
                     /protein_id="AAG49779.1"
                     /db_xref="GI:12247202"
                     /translation="AEKEVTRMVIIMVVFFLICWVPYAYVAFYIFTHQGSEFGPIFMT
                     VPAFFAKSSAIYNP"
     exon            <1. .>175
                     /number=4
BASE COUNT       34 a     55 c     34 g     52 t
ORIGIN

  Query Match             82.5%;  Score 19.8;  DB 5;  Length 175;
  Best Local Similarity   91.3%;  Pred. No. 37;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           |||||||||||||||||| | ||
Db     140 GCTTTCTTTGCCAAGAGCTCTGC 162


Search completed: March 17, 2003, 11:25:45
Job time : 313.903 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:42:12 ; Search time 137.548 Seconds
(without alignments)
392.938 Million cell updates/sec

Title: US-09-836-439-3
Perfect score: 24
Sequence: 1 gctttctttgccaagagcgccgca 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21.4	89.2	3016	17	AAT17115	Rhodopsin gene. H
	2	21.4	89.2	6953	17	AAT17116	Rhodopsin gene. H
	3	18.8	78.3	8641	20	AAX84344	Stealth virus nucl
	4	18.2	75.8	3129	14	AAQ43543	Rhodopsin gene. H
	5	18.2	75.8	3190	23	ABL26876	Drosophila melanog
c	6	17.6	73.3	1563	21	AAZ53713	Neisseria gonorrhe
	7	17.4	72.5	2209	22	AAS41923	Genomic sequence #
c	8	17.4	72.5	2601	20	AAX90036	Fervidobacterium p
c	9	17	70.8	36135	22	AAK84218	Human immune/haema
	10	16.8	70.0	467	23	ABV12459	Human prostate exp
	11	16.8	70.0	506	23	ABV33599	Human prostate exp
	12	16.8	70.0	1679	24	ABK63732	Rat sequence diffe
	13	16.8	70.0	2304	23	ABL02991	Drosophila melanog
	14	16.8	70.0	4446	23	ABL02990	Drosophila melanog
	15	16.6	69.2	408	24	ABK80729	Bacillus clausii g
	16	16.6	69.2	408	24	ABK80749	Bacillus clausii g
	17	16.6	69.2	740	22	AAK59608	Human immune/haema
c	18	16.6	69.2	795	24	ABN66622	Streptococcus poly
c	19	16.6	69.2	1845	17	AAT33631	Human p65 oncofoet
	20	16.6	69.2	4785	12	AAQ14072	Cyclodextrin gluca
	21	16.6	69.2	4785	12	AAQ14942	Cyclodextrin gluca
	22	16.6	69.2	4785	14	AAQ48788	B.ohbensis mutant
	23	16.6	69.2	4785	14	AAQ48794	B.ohbensis Ala(188
	24	16.6	69.2	4785	14	AAQ48795	B.ohbensis Ser(188
	25	16.6	69.2	4785	14	AAQ48796	B.ohbensis Ser(188
	26	16.6	69.2	4785	14	AAQ48797	B.ohbensis Trp(188
c	27	16.6	69.2	34185	21	AAC62130	Nucleotide sequenc
c	28	16.6	69.2	240825	22	AAF24497	Human PG-3 gene.
c	29	16.4	68.3	336	23	ABV03802	Human prostate exp
c	30	16.4	68.3	337	23	ABV12971	Human prostate exp
c	31	16.4	68.3	385	23	ABV34093	Human prostate exp
c	32	16.4	68.3	385	23	ABV42957	Human prostate exp
c	33	16.4	68.3	433	21	AAA44043	Human secreted exp
c	34	16.4	68.3	956	22	ABA12912	Human nervous syst
c	35	16.4	68.3	2013	22	ABA18522	Human nervous syst
c	36	16.4	68.3	2060	23	ABV22500	Human prostate exp
c	37	16.4	68.3	2060	23	ABV28314	Human prostate exp
c	38	16.4	68.3	5801	21	AAZ93310	Sequence encoding
c	39	16.2	67.5	392	21	AAA57072	Human colon cancer
c	40	16.2	67.5	395	23	ABV06135	Human prostate exp
c	41	16.2	67.5	459	22	AAL13938	Human breast cance
	42	16.2	67.5	485	22	AAH84028	Eulemar fulvus olf
c	43	16.2	67.5	493	21	AAA57090	Human colon cancer
c	44	16.2	67.5	563	24	ABN73103	Bovine embryonic g
c	45	16.2	67.5	586	22	AAL22800	Human breast cance

ALIGNMENTS

RESULT 1
AAT17115
ID AAT17115 standard; DNA; 3016 BP.
XX
AC AAT17115;
XX
DT 06-JUL-1996 (first entry)
XX
DE Rhodopsin gene.
XX
KW Human; rhodopsin; transversion; mutation; retinitis pigmentosa;
KW probe; primer; hybridisation; polymerase chain reaction; PCR;
KW eye; rod; retina; diagnostic; prenatal diagnosis; photoreceptor; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 5'UTR 200..294
FT /*tag= a
FT 5'UTR 202..294
FT /*tag= b
FT /note= "Alternative 5'-UTR"
FT primer_bind complement (231..250)
FT /*tag= c
FT /note= "Binds primer 348 (AAT17121)"
FT CDS 295..5278
FT /*tag= d
FT /product= Rhodopsin
FT misc_binding complement (354..372)
FT /*tag= e
FT /note= "Binds probes AAT17117 (mutant) and AAT17118 (normal)"
FT

```
FT   misc_binding    complement (355..369)
FT                   /*tag=  f
FT                   /note= "Binds probes AAT17119 (mutant) and AAT17120
FT                          (normal)"
FT   mutation        362
FT                   /*tag=  g
FT                   /note= "Substitution with A in mutant sequence"
FT   primer_bind     362..381
FT                   /*tag=  h
FT                   /note= "Binds primers 485 (AAT17122) (normal) and 502
FT                          (mutant)"
XX
PN   US5498521-A.
XX
PD   12-MAR-1996.
XX
PF   24-JAN-1990;   90US-0469215.
XX
PR   11-MAR-1993;   93US-0033081.
PR   24-JAN-1990;   90US-0469215.
PR   11-DEC-1991;   91US-0805123.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Berson EL,  Dryja TP;
XX
DR   WPI; 1996-159684/16.
DR   P-PSDB; AAR93116.
XX
PT   Diagnosis of hereditary retinal degenerative diseases e.g. retinitis
PT   pigmentosa, - caused by a human photoreceptor protein mutation, by
PT   detection of the mutation by PCR amplification or hybridisation
PT   methods
XX
PS   Example 1; Column 19-24; 71pp; English.
XX
CC   This sequence encodes human rhodopsin, and is shown without
CC   introns.  The full sequence, with introns, is shown in AAT17116.
CC   Substitution of histidine for the normal nonpolar amino acid
CC   proline at position 23, by substitution of C with A in codon-23,
CC   results in a dysfunctional or absent molecule, affecting rod
CC   function, and is linked with autosomal dominant retinitis
CC   pigmentosa.  Probes AAT17117 and AAT17119 bind to the C-to-A
CC   transversion mutation sequence, and probes AAT17118 and AAT17120 bind
CC   to the corresponding normal sequence.  Primers 485 (AAT17122) and 502
CC   (AAT17123) may be used along with primer 348 (AAT17121) to amplify
CC   mutant and normal sequences, respectively, by PCR.  Mutations in the
CC   retinal degeneration slow protein and retinal rod
CC   cGMP-phosphodiesterase genes are also implicated in retinitis
CC   pigmentosa.  Detection of any of these mutations in a foetus or
CC   patient may be used in diagnosis.
XX
SQ   Sequence 3016 BP; 689 A; 863 C; 753 G; 711 T; 0 other;

  Query Match             89.2%;  Score 21.4;  DB 17;  Length 3016;
  Best Local Similarity   95.7%;  Pred. No. 1.7;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || ||||||||||||||||||||
Db    1168 GCGTTCTTTGCCAAGAGCGCCGC 1190


RESULT 2
AAT17116
ID   AAT17116 standard; DNA; 6953 BP.
XX
AC   AAT17116;
XX
DT   06-JUL-1996  (first entry)
XX
DE   Rhodopsin gene.
XX
KW   Human; rhodopsin; transversion; mutation; retinitis pigmentosa;
KW   intron; probe; primer; hybridisation; polymerase chain reaction; PCR;
KW   eye; rod; retina; diagnostic; prenatal diagnosis; photoreceptor; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           200..294
FT                   /*tag=  a
FT   5'UTR           202..294
FT                   /*tag=  b
FT                   /note= "Alternative 5'-UTR"
FT   primer_bind     complement (231..250)
FT                   /*tag=  c
FT                   /note= "Binds primer 348 (AAT17121)"
FT   exon            295..655
FT                   /*tag=  d
FT                   /number= 1
FT   misc_binding    complement (354..372)
FT                   /*tag=  e
FT                   /note= "Binds probes AAT17117 (mutant) and AAT17118
FT                          (normal)"
FT   misc_binding    complement (355..369)
FT                   /*tag=  f
FT                   /note= "Binds probes AAT17119 (mutant) and AAT17120
FT                          (normal)"
FT   mutation        362
FT                   /*tag=  g
FT                   /note= "Substitution with A in mutant sequence"
FT   primer_bind     362..381
FT                   /*tag=  h
FT                   /note= "Binds primers 485 (AAT17122) (normal) and 502
FT                          (mutant)"
FT   intron          666..2438
FT                   /*tag=  i
FT                   /number= 1
FT   exon            2439..2607
FT                   /*tag=  j
FT                   /number= 2
FT   intron          2608..3812
FT                   /*tag=  k
FT                   /number= 2
FT   exon            3813..3978
FT                   /*tag=  l
FT                   /number= 3
FT   intron          3979..4094
FT                   /*tag=  m
FT                   /number= 3
FT   exon            4095..4334
FT                   /*tag=  n
FT                   /number= 4
FT   intron          4335..5167
FT                   /*tag=  o
FT                   /number= 4
FT   exon            5168..5278
FT                   /*tag=  p
FT                   /number= 5
XX
PN   US5498521-A.
XX
PD   12-MAR-1996.
XX
PF   24-JAN-1990;   90US-0469215.
XX
PR   11-MAR-1993;   93US-0033081.
PR   24-JAN-1990;   90US-0469215.
PR   11-DEC-1991;   91US-0805123.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Berson EL,  Dryja TP;
XX
DR   WPI; 1996-159684/16.
```

DR P-PSDB; AAR93116.
XX
PT Diagnosis of hereditary retinal degenerative diseases e.g. retinitis
PT pigmentosa, - caused by a human photoreceptor protein mutation, by
PT detection of the mutation by PCR amplification or hybridisation
PT methods
XX
PS Example 1; Column 23-30; 71pp; English.
XX
CC This sequence encodes human rhodopsin, and is shown in full with
CC introns. The corresponding sequence without introns is shown in
CC AAT17116. Substitution of histidine for the normal nonpolar amino
CC acid proline at position 23, by substitution of C with A in
CC codon-23, results in a dysfunctional or absent molecule, affecting
CC rod function, and is linked with autosomal dominant retinitis
CC pigmentosa. Probes AAT17117 and AAT17119 bind to the C-to-A
CC transversion mutation sequence, and probes AAT17118 and AAT17120 bind
CC to the corresponding normal sequence. Primers 485 (AAT17122) and 502
CC (AAT17123) may be used along with primer 348 (AAT17121) to amplify
CC mutant and normal sequences, respectively, by PCR. Mutations in the
CC retinal degeneration slow protein and retinal rod
CC cGMP-phosphodiesterase genes are also implicated in retinitis
CC pigmentosa. Detection of any of these mutations in a foetus or
CC patient may be used in diagnosis.
XX
SQ Sequence 6953 BP; 1523 A; 2022 C; 1797 G; 1611 T; 0 other;

Query Match 89.2%; Score 21.4; DB 17; Length 6953;
Best Local Similarity 95.7%; Pred. No. 2;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || ||||||||||||||||||||
Db    4272 GCGTTCTTTGCCAAGAGCGCCGC 4294
```

RESULT 3
AAX84344
ID AAX84344 standard; DNA; 8641 BP.
XX
AC AAX84344;
XX
DT 08-SEP-1999 (first entry)
XX
DE Stealth virus nucleic acid clone, SEQ ID NO: 36.
XX
KW Stealth virus; detection; diagnosis; infection; ss.
XX
OS Stealth virus.
XX
PN WO9934019-A1.
XX
PD 08-JUL-1999.
XX
PF 30-DEC-1998; 98WO-US27744.
XX
PR 30-DEC-1997; 97US-0001184.
XX
PA (MART/) MARTIN W J.
XX
PI Martin WJ;
XX
DR WPI; 1999-405521/34.
XX
PT Novel strains of stealth virus
XX
PS Claim 19; Page 76-79; 95pp; English.
XX
CC This sequence represents a Stealth virus nucleic acid clone. The
CC invention relates to a method of detecting and characterising a stealth
CC virus by reacting a sample suspected of containing a stealth virus with a
CC probe from a known stealth virus and sequencing the resultant isolated
CC nucleotide. The method comprises the steps of: (a) isolating DNA or RNA
CC from a sample suspected of containing a stealth virus, e.g. a culture of
CC cells showing a viral cytopathic effect; (b) testing the reactivity of
CC the isolated DNA or RNA with a molecular probe that contains at least 18
CC or more contiguous nucleotides identical to sequence previously
CC identified from a stealth virus; and, optionally (c) sequencing the
CC isolated DNA or RNA molecules that react with the probe. The method is
CC used to detect stealth virus in a biological product, food or in the
CC environment. The method is also used to evaluate agents for their
CC inhibitory or stimulatory effects on stealth virus replication and to
CC determine capacity of the virus to recombine with and potentially alter
CC the nucleic acid sequences of a cell or bacterium.
XX
SQ Sequence 8641 BP; 2101 A; 2031 C; 2018 G; 2476 T; 15 other;

Query Match 78.3%; Score 18.8; DB 20; Length 8641;
Best Local Similarity 90.9%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       2 CTTTCTTTGCCAAGAGCGCCGC 23
           ||||||||  ||||||||||||
Db    8149 CTTTCTTTCGCAAGAGCGCCGC 8170
```

RESULT 4
AAQ43543
ID AAQ43543 standard; cDNA; 3129 BP.
XX
AC AAQ43543;
XX
DT 11-NOV-1993 (first entry)
XX
DE Rhodopsin gene.
XX
KW Human; rhodopsin; mutant; retinal degeneration; primer; probe;
KW hereditary; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT prim_transcript 200..1341
FT /*tag= a
FT CDS 295..1341
FT /*tag= b
XX
PN WO9312134-A.
XX
PD 24-JUN-1993.
XX
PF 08-DEC-1992; 92WO-US10536.
XX
PR 11-DEC-1991; 91US-0805123.
XX
PA (HARD) HARVARD COLLEGE.
XX
PI Berson EL, Dryja TP;
XX
DR WPI; 1993-214088/26.
DR P-PSDB; AAR38483.
XX
PT Probe or primer contg. sequence of human retinal degeneration
PT slow protein mutant - used to diagnose hereditary retinal
PT degenerative diseases
XX
PS Disclosure; Fig 1; 56pp; English.
XX
CC The sequence given represents the human rhodopsin cDNA. Mutant
CC versions of this sequence encode proteins which cause retinal
CC degeneration. These sequences may be identified using primers/
CC probes described in the invention (see also AAQ43545-48) and may be
CC used to diagnose hereditary retinal degeneration. This sequence is
CC the closest approximation to the gene sequence as the sequence given
CC in the specification is not printed clearly.
XX

```
SQ   Sequence 3129 BP; 694 A; 999 C; 654 G; 735 T; 47 other;

  Query Match             75.8%;  Score 18.2;  DB 14;  Length 3129;
  Best Local Similarity   87.0%;  Pred. No. 61;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || |||||||||||||  ||||||
Db    1168 GCGTTCTTTGCCAAGTCCGCCGC 1190


RESULT 5
ABL26876
ID   ABL26876 standard; DNA; 3190 BP.
XX
AC   ABL26876;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 32101.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 32101; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3190 BP; 831 A; 719 C; 649 G; 991 T; 0 other;

  Query Match             75.8%;  Score 18.2;  DB 23;  Length 3190;
  Best Local Similarity   87.0%;  Pred. No. 61;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 CTTTCTTTGCCAAGAGCGCCGCA 24
           ||||||| |||||||| || |||
Db    2504 CTTTCTTGGCCAAGAGTGCAGCA 2526


RESULT 6
AAZ53713/c
ID   AAZ53713 standard; DNA; 1563 BP.
XX
AC   AAZ53713;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 506 partial DNA sequence SEQ ID NO:1375.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   P-PSDB; AAY74951.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 7; Page 748; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence 1563 BP; 287 A; 372 C; 487 G; 417 T; 0 other;

  Query Match             73.3%;  Score 17.6;  DB 21;  Length 1563;
  Best Local Similarity   83.3%;  Pred. No. 1.1e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGCA 24
           ||||| |||| ||| | |||||||
Db      63 GCTTTGTTTGACAACACCGCCGCA 40


RESULT 7
AAS41923
ID   AAS41923 standard; DNA; 2209 BP.
XX
AC   AAS41923;
```

XX
DT 17-DEC-2001 (first entry)
XX
DE Genomic sequence #239 encoding novel human enzyme polypeptide.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotropic; anticoagulant; ds.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465566/50.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases
XX
PS Disclosure; SEQ ID No 2049; 1180pp; English.
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAS41685-AAS42192 represent DNA sequences encoding for the novel human
CC enzyme polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2209 BP; 629 A; 412 C; 442 G; 726 T; 0 other;

Query Match 72.5%; Score 17.4; DB 22; Length 2209;
Best Local Similarity 94.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCTTTCTTTGCCAAGAGCG 19
||||||||||||||||| |
Db 843 GCTTTCTTTGCCAAGAGTG 861

RESULT 8
AAX90036/c
ID AAX90036 standard; DNA; 2601 BP.
XX
AC AAX90036;
XX
DT 17-SEP-1999 (first entry)
XX
DE Fervidobacterium pennavorans Ven 5 thermostable pullulanase DNA.
XX
KW Fervidobacterium pennavorans; Ven 5; thermostable pullulanase; pulA;
KW industrial saccharification; ss.
XX
OS Fervidobacterium pennavorans.
XX
PN WO9935274-A2.
XX
PD 15-JUL-1999.
XX
PF 05-JAN-1999; 99WO-IB00069.
XX
PR 07-JAN-1998; 98US-0003834.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Duffner F, Jorgensen PL;
XX
DR WPI; 1999-430401/36.
DR P-PSDB; AAY24380.
XX
PT Nucleic acid encoding Fervidobacterium sp. Ven 5 thermostable
PT pullulanase, useful for industrial saccharification processes
XX
PS Claim 1; Page 40-41; 52pp; English.
XX
CC The present sequence encodes Fervidobacterium pennavorans Ven 5
CC thermostable pullulanase. Thermostable pullulanases are useful in
CC industrial saccharification processes (see EP63909).
XX
SQ Sequence 2601 BP; 881 A; 494 C; 596 G; 630 T; 0 other;

Query Match 72.5%; Score 17.4; DB 20; Length 2601;
Best Local Similarity 94.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TTTCTTTGCCAAGAGCGCC 21
|||||||| ||||||||||
Db 1924 TTTCTTTGGCAAGAGCGCC 1906

RESULT 9
AAK84218/c
ID AAK84218 standard; DNA; 36135 BP.
XX
AC AAK84218;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39030.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 39030; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.

```
XX
SQ   Sequence 36135 BP; 10321 A; 8075 C; 8063 G; 9676 T; 0 other;

  Query Match             70.8%;  Score 17;  DB 22;  Length 36135;
  Best Local Similarity   100.0%;  Pred. No. 3.5e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 CTTTCTTTGCCAAGAGC 18
           |||||||||||||||||
Db     920 CTTTCTTTGCCAAGAGC 904


RESULT 10
ABV12459
ID   ABV12459 standard; cDNA; 467 BP.
XX
AC   ABV12459;
XX
DT   13-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 12450.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS   Claim 1; Page 2053; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 467 BP; 71 A; 115 C; 132 G; 149 T; 0 other;

  Query Match             70.0%;  Score 16.8;  DB 23;  Length 467;
  Best Local Similarity   90.0%;  Pred. No. 2.1e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGC 20
           |||||||||| ||| |||||
Db     330 GCTTTCTTTGGCAAAAGCGC 349


RESULT 11
ABV33599
ID   ABV33599 standard; cDNA; 506 BP.
XX
AC   ABV33599;
XX
DT   16-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 33590.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS   Claim 1; Page 7105; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 506 BP; 80 A; 129 C; 145 G; 152 T; 0 other;

  Query Match             70.0%;  Score 16.8;  DB 23;  Length 506;
  Best Local Similarity   90.0%;  Pred. No. 2.1e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGC 20
           |||||||||| ||| |||||
Db     376 GCTTTCTTTGGCAAAAGCGC 395
```

```
RESULT 12
ABK63732
ID   ABK63732 standard; cDNA; 1679 BP.
XX
AC   ABK63732;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Rat sequence differentially expressed in response to a hepatotoxin #1639.
XX
KW   Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
KW   differential expression; centrilobular necrosis; steatosis.
XX
OS   Rattus norvegicus.
XX
PN   WO200210453-A2.
XX
PD   07-FEB-2002.
XX
PF   30-JUL-2001; 2001WO-US23872.
XX
PR   31-JUL-2000; 2000US-222040P.
PR   02-NOV-2000; 2000US-244880P.
PR   11-MAY-2001; 2001US-290029P.
PR   15-MAY-2001; 2001US-290645P.
PR   22-MAY-2001; 2001US-292336P.
PR   06-JUN-2001; 2001US-295798P.
PR   13-JUN-2001; 2001US-297457P.
PR   19-JUN-2001; 2001US-298884P.
PR   09-JUL-2001; 2001US-303459P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Mendrick D,  Porter MW,  Johnson KR,  Castle AL,  Elashoff MR;
XX
DR   WPI; 2002-241625/29.
XX
PT   Predicting toxic effects of compounds or the progression of these toxic
PT   effects by determining the changes in gene expression in tissues or
PT   cells exposed to the toxin and comparing these to gene expression in
PT   unexposed tissues or cells -
XX
PS   Claim 1; Seq ID No 1639; 239pp; English.
XX
CC   The invention relates to methods for predicting toxic effects of
CC   compounds or the progression of these toxic effects by determining the
CC   global changes in gene expression in tissues or cells exposed to the
CC   toxin and comparing these to gene expression in unexposed tissues or
CC   cells. Also included are methods of predicting at least one toxic
CC   effect of a compound or progression of a toxic effect, preferably the
CC   hepatotoxicity of a compound, comprising detecting the level of
CC   expression in a tissue or cell sample exposed to the compound of two or
CC   more genes listed in the specification, where  differential expression of
CC   the genes is indicative of at least one toxic effect or progression.
CC   The method can also be used to identify an agent which modulates the
CC   toxic response and predict cellular pathways that a compound modulates
CC   in a cell. The methods utilise a set of at least two probes (on a solid
CC   support in kit form), where each of the probes comprises a sequence that
CC   specifically hybridises to a gene listed in the specification, a computer
CC   system comprising a database containing information identifying the
CC   expression level in a tissue or cell sample exposed to a hepatotoxin of a
CC   set of genes comprising at least two genes listed in the specification,
CC   and a user interface to view the information used to present information
CC   identifying the expression level in a tissue or cell of at least one gene
CC   listed in the specification. The method is useful for elucidating global
CC   changes in gene expression and for identifying toxicity markers in
CC   tissues or cell exposed to a known toxin. The genes may be used as
CC   toxicity markers in drug screening and toxicity assays. The genes and
CC   gene expression information may be used as diagnostic markers for the
CC   prediction or identification of the physiological state of tissue or cell
CC   sample that has been exposed to a compound or agent. Hepatotoxicity
CC   is characterised by centrilobular necrosis and steatosis. The present
CC   sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC   which is differentially expressed in response to a hepatotoxic agent.
XX
SQ   Sequence 1679 BP; 377 A; 471 C; 483 G; 348 T; 0 other;

  Query Match             70.0%;  Score 16.8;  DB 24;  Length 1679;
  Best Local Similarity   90.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 TCTTTGCCAAGAGCGCCGCA 24
           ||||||||  ||||||||||
Db      61 TCTTTGCCTTGAGCGCCGCA 80


RESULT 13
ABL02991
ID   ABL02991 standard; cDNA; 2304 BP.
XX
AC   ABL02991;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 3455.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB58888.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 3455; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2304 BP; 541 A; 640 C; 721 G; 402 T; 0 other;

  Query Match             70.0%;  Score 16.8;  DB 23;  Length 2304;
  Best Local Similarity   90.0%;  Pred. No. 2.7e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 TTCTTTGCCAAGAGCGCCGC 23
           |||||||||  |||||||  ||
Db     226 TTCTTTGCCCAGAGCGCAGC 245
```

```
RESULT 14
ABL02990
ID   ABL02990 standard; cDNA; 4446 BP.
XX
AC   ABL02990;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 3452.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB58887.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 3452; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4446 BP; 1191 A; 1023 C; 1084 G; 1148 T; 0 other;

  Query Match           70.0%;  Score 16.8;  DB 23;  Length 4446;
  Best Local Similarity  90.0%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 TTCTTTGCCAAGAGCGCCGC 23
           ||||||||| ||||||| ||
Db     613 TTCTTTGCCCAGAGCGCAGC 632


RESULT 15
ABK80729
ID   ABK80729 standard; DNA; 408 BP.
XX
AC   ABK80729;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Bacillus clausii genomic sequence tag (GST) #3572.
XX
KW   Differential gene expression; genomic sequenced tag; GST;
KW   altered culture condition; environmental stress;
KW   physiological provocation; ds.
XX
OS   Bacillus clausii.
XX
PN   WO200229113-A2.
XX
PD   11-APR-2002.
XX
PF   05-OCT-2001; 2001WO-US31437.
XX
PR   06-OCT-2000; 2000US-0680598.
PR   27-MAR-2001; 2001US-279526P.
XX
PA   (NOVO ) NOVOZYMES BIOTECH INC.
PA   (NOVO ) NOVOZYMES AS.
XX
PI   Berka R,  Clausen IG;
XX
DR   WPI; 2002-416684/44.
XX
PT   Monitoring differential expression of several genes in first Bacillus
PT   cell relative to expression of same genes in one or more second
PT   Bacillus cells, by using substrate containing Bacillus genomic
PT   sequenced tag array -
XX
PS   Claim 11; SEQ ID NO 8020; 200pp; English.
XX
CC   The invention describes a method of monitoring differential expression of
CC   genes in a first Bacillus cell relative to expression of the genes in
CC   other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC   isolated from Bacillus cells to a substrate containing array of Bacillus
CC   genomic sequenced tags (GST), examining the array, and determining
CC   relative gene expression by an observed hybridisation reporter signal of
CC   a spot in the array. The method is useful for measuring the expression of
CC   genes in a first Bacillus cell relative to expression of the same genes
CC   in one or more second Bacillus cells. The method is useful for monitoring
CC   global expression of several genes from a Bacillus cell, discovering new
CC   genes, identifying possible functions of unknown open reading frames and
CC   monitoring gene copy number variation and stability. Monitoring changes
CC   in expression of genes may be used to provide a representation of the way
CC   in which Bacillus cells adapt to changes in culture conditions,
CC   environmental stress or other physiological provocation. Extensive
CC   follow-up characterisation is unnecessary, when one spot on an array
CC   equals one gene or one open reading frame, since sequence information is
CC   available. This sequence represents a genomic sequence tag (GST) used in
CC   the method of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 408 BP; 106 A; 86 C; 104 G; 111 T; 1 other;

  Query Match           69.2%;  Score 16.6;  DB 24;  Length 408;
  Best Local Similarity  82.6%;  Pred. No. 2.5e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || |||||  ||||| |||||||
Db     374 GCATTCTTGCCCAAGTGCGCCGC 396


Search completed: March 17, 2003, 10:50:45
Job time : 146.715 secs
```

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 10:23:12 ; Search time 848.387 Seconds
(without alignments)
458.154 Million cell updates/sec

Title: US-09-836-439-3
Perfect score: 24
Sequence: 1 gctttctttgccaagagcgccgca 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21.4	89.2	303	14	BM694379	BM694379 UI-E-CI1-
	2	21.4	89.2	319	14	BQ637565	BQ637565 he11g04.y
	3	21.4	89.2	337	14	BM723222	BM723222 UI-E-EJ0-
c	4	21.4	89.2	340	14	BM682444	BM682444 UI-E-EJ0-
	5	21.4	89.2	349	14	BM703950	BM703950 UI-E-CK1-
	6	21.4	89.2	441	14	BM690151	BM690151 UI-E-CL0-
	7	21.4	89.2	442	9	AL712412	AL712412 DKFZp686O
	8	21.4	89.2	446	14	BM696193	BM696193 UI-E-CL1-
	9	21.4	89.2	469	14	BM688227	BM688227 UI-E-CL0-
	10	21.4	89.2	471	9	AL712402	AL712402 DKFZp686N
	11	21.4	89.2	471	14	BQ250368	BQ250368 TaE25006A
	12	21.4	89.2	474	14	BM688069	BM688069 UI-E-CL0-
c	13	21.4	89.2	478	13	BM662914	BM662914 UI-E-CK0-
	14	21.4	89.2	493	14	BM690137	BM690137 UI-E-CL0-
	15	21.4	89.2	493	14	BM691592	BM691592 UI-E-CI1-
	16	21.4	89.2	496	14	BM690311	BM690311 UI-E-CL0-
	17	21.4	89.2	499	14	BQ640665	BQ640665 he31f09.y
	18	21.4	89.2	503	14	BQ636862	BQ636862 he02e01.y
	19	21.4	89.2	504	14	BM686188	BM686188 UI-E-CK0-
	20	21.4	89.2	505	14	T27877	T27877 EST19137 Hu
	21	21.4	89.2	510	14	BQ637734	BQ637734 he13h04.y
	22	21.4	89.2	513	14	BM690319	BM690319 UI-E-CL0-
	23	21.4	89.2	514	14	BQ640660	BQ640660 he31f02.y
	24	21.4	89.2	526	14	BQ639829	BQ639829 he20h01.y
	25	21.4	89.2	530	14	BM690115	BM690115 UI-E-CL0-
	26	21.4	89.2	530	14	BQ638124	BQ638124 hd18g07.y
	27	21.4	89.2	530	14	BQ638268	BQ638268 hd20e05.y
	28	21.4	89.2	535	14	BQ640241	BQ640241 he26a01.y
	29	21.4	89.2	549	14	BQ636918	BQ636918 he03c02.y
	30	21.4	89.2	549	14	BQ639356	BQ639356 he14e06.y
	31	21.4	89.2	562	14	BQ636602	BQ636602 hd11h09.y
	32	21.4	89.2	575	14	BQ640490	BQ640490 he29d06.y
	33	21.4	89.2	576	14	BM704430	BM704430 UI-E-CI1-
	34	21.4	89.2	578	14	BM685757	BM685757 UI-E-CK0-
	35	21.4	89.2	578	14	BM688176	BM688176 UI-E-CL0-
	36	21.4	89.2	580	14	BM685912	BM685912 UI-E-CK0-
	37	21.4	89.2	587	14	BQ637208	BQ637208 he07a10.y
	38	21.4	89.2	590	14	BQ640394	BQ640394 he27h12.y
	39	21.4	89.2	591	9	AL711507	AL711507 DKFZp686D
	40	21.4	89.2	595	14	BQ638291	BQ638291 hd20g07.y
	41	21.4	89.2	599	14	BQ636906	BQ636906 he03a11.y
	42	21.4	89.2	604	14	BQ638120	BQ638120 hd18g02.y
	43	21.4	89.2	605	9	AL712251	AL712251 DKFZp686L
	44	21.4	89.2	606	14	BQ638554	BQ638554 hd24h07.y
	45	21.4	89.2	606	14	BQ639637	BQ639637 he18a09.y

ALIGNMENTS

RESULT 1
BM694379
LOCUS BM694379 303 bp mRNA linear EST 28-FEB-2002
DEFINITION UI-E-CI1-afp-p-18-0-UI.r1 UI-E-CI1 Homo sapiens cDNA clone UI-E-CI1-afp-p-18-0-UI 5', mRNA sequence.
ACCESSION BM694379
VERSION BM694379.1 GI:19007637
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 303)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa

Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 Reverse.

```
FEATURES             Location/Qualifiers
     source          1. .303
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-E-CI1-afp-p-18-0-UI"
                     /clone_lib="UI-E-CI1"
                     /tissue_type="RPE and Choroid"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-CI1 is a normalized cDNA library containing the
                     following tissue(s): RPE and Choroid. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is ACCTA.
                     This library was created for the program, Gene Discovery
                     in the Visual System, supported by National Eye Institute
                     (NEI)."
BASE COUNT       70 a    101 c     67 g     65 t
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 14;  Length 303;
  Best Local Similarity   95.7%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || ||||||||||||||||||||
Db     171 GCGTTCTTTGCCAAGAGCGCCGC 193


RESULT 2
BQ637565
LOCUS       BQ637565                319 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  he11g04.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he
            Homo sapiens cDNA clone he11g04 5', mRNA sequence.
ACCESSION   BQ637565
VERSION     BQ637565.1  GI:21762024
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 319)
  AUTHORS   Wistow,G., Bernstein,S.L., Wyatt,M.K., Ray,S., Behal,A., Touchman
            ,J.W., Bouffard,G., Smith,D. and Peterson,K.
  TITLE     Expressed sequence tag analysis of human retina for the NEIBank
            Project: Retbindin, an abundant, novel retinal cDNA and alternative
            splicing of other retina-preferred gene transcripts
  JOURNAL   Mol. Vis. 8 (4),  (2002) In press
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 11  row: g  column: 04
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .319
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="he11g04"
                     /clone_lib="Human Retina cDNA (Un-normalized, unamplified
                     ): hd/he"
                     /tissue_type="Retina"
                     /dev_stage="Adult"
                     /lab_host="EMDH10B"
                     /note="Organ: Eye; Vector: pSPORT1; Neural retina tissue
                     was dissected from two 80 year old donors with no observed
                     eye disease. 100ug of total RNA was used for library
                     construction. A directionally cloned cDNA library in the
                     pSPORT1 vector (Life Technologies) was constructed at
                     Bioserve Biotechnology (Laurel MD) essentially following
                     the protocols of the SuperScript Plasmid System full
                     details of which are contained in the manufacturer's
                     Instruction manual (http://www.lifetech.com/). First
                     strand synthesis was carried out using a Not I
                     primer-adapter [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3'
                     ]. EST analysis was performed on the unamplified library
                     at the NIH Intramural Sequencing Center (NISC)."
BASE COUNT       72 a    108 c     75 g     64 t
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 14;  Length 319;
  Best Local Similarity   95.7%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || ||||||||||||||||||||
Db     204 GCGTTCTTTGCCAAGAGCGCCGC 226


RESULT 3
BM723222
LOCUS       BM723222                337 bp    mRNA    linear   EST 01-MAR-2002
DEFINITION  UI-E-EJ0-aio-i-15-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone
            UI-E-EJ0-aio-i-15-0-UI 5', mRNA sequence.
ACCESSION   BM723222
VERSION     BM723222.1  GI:19044349
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 337)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 295-332, >AT_rich#Low_complexity (matched compliment)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .337
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-E-EJ0-aio-i-15-0-UI"
                     /clone_lib="UI-E-EJ0"
                     /tissue_type="fetal eyes, lens, eye anterior segment,
                     optic nerve, retina, Retina Foveal and Macular, RPE and
                     Choroid"
```

/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-EJ0 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA ; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT; optic nerve, CCATTAAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."
BASE COUNT 86 a 105 c 74 g 72 t
ORIGIN

Query Match 89.2%; Score 21.4; DB 14; Length 337;
Best Local Similarity 95.7%; Pred. No. 24;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 GCTTTCTTTGCCAAGAGCGCCGC 23
          || ||||||||||||||||||||
Db      1 GCGTTCTTTGCCAAGAGCGCCGC 23
```

RESULT 4
BM682444/c
LOCUS BM682444 340 bp mRNA linear EST 27-FEB-2002
DEFINITION UI-E-EJ0-aio-i-15-0-UI.s1 UI-E-EJ0 Homo sapiens cDNA clone UI-E-EJ0-aio-i-15-0-UI 3', mRNA sequence.
ACCESSION BM682444
VERSION BM682444.1 GI:18992340
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 340)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA sequence: 1-37, >AT_rich#Low_complexity (matched compliment)
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .340
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-EJ0-aio-i-15-0-UI"
/clone_lib="UI-E-EJ0"
/tissue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-EJ0 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA ; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT; optic nerve, CCATTAAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).
TAG_LIB=UI-E-EJ0
TAG_TISSUE=human retina
TAG_SEQ=CCGCG"
BASE COUNT 71 a 76 c 105 g 88 t
ORIGIN

Query Match 89.2%; Score 21.4; DB 14; Length 340;
Best Local Similarity 95.7%; Pred. No. 24;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 GCTTTCTTTGCCAAGAGCGCCGC 23
          || ||||||||||||||||||||
Db    331 GCGTTCTTTGCCAAGAGCGCCGC 309
```

RESULT 5
BM703950
LOCUS BM703950 349 bp mRNA linear EST 28-FEB-2002
DEFINITION UI-E-CK1-afk-m-09-0-UI.r1 UI-E-CK1 Homo sapiens cDNA clone UI-E-CK1-afk-m-09-0-UI 5', mRNA sequence.
ACCESSION BM703950
VERSION BM703950.1 GI:19017208
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 349)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .349
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-CK1-afk-m-09-0-UI"

/clone_lib="UI-E-CK1"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-CK1 is a normalized cDNA library containing the following tissue(s): Retina Foveal and Macular. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTCC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 96 a 116 c 76 g 61 t
ORIGIN

Query Match 89.2%; Score 21.4; DB 14; Length 349;
Best Local Similarity 95.7%; Pred. No. 24;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || |||||||||||||||||||||
Db      37 GCGTTCTTTGCCAAGAGCGCCGC 59
```

RESULT 6
BM690151
LOCUS BM690151 441 bp mRNA linear EST 28-FEB-2002
DEFINITION UI-E-CL0-acd-f-11-0-UI.r1 UI-E-CL0 Homo sapiens cDNA clone UI-E-CL0-acd-f-11-0-UI 5', mRNA sequence.
ACCESSION BM690151
VERSION BM690151.1 GI:19003409
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 441)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-CL0-acd-f-11-0-UI"
/clone_lib="UI-E-CL0"
/tissue_type="human retina"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-CL0 is a cDNA library containing the following tissue(s): retina. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCGCG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 97 a 143 c 100 g 101 t
ORIGIN

Query Match 89.2%; Score 21.4; DB 14; Length 441;
Best Local Similarity 95.7%; Pred. No. 26;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || |||||||||||||||||||||
Db     301 GCGTTCTTTGCCAAGAGCGCCGC 323
```

RESULT 7
AL712412
LOCUS AL712412 442 bp mRNA linear EST 22-MAR-2002
DEFINITION DKFZp686O1888_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone DKFZp686O1888 5', mRNA sequence.
ACCESSION AL712412
VERSION AL712412.1 GI:19695767
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 442)
AUTHORS Wambutt,R., Heubner,D., Mewes,W., Weil,B. and Wiemann,S.
TITLE EST (Wambutt,R., Heubner,D., Mewes,H.W., Weil,B. and Wiemann,S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Wambutt R
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
No s1 sequence available.
This clone (DKFZp686O1888) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES Location/Qualifiers
source 1. .442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686O1888"
/clone_lib="686 (synonym: hlcc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB; cDNA-collection"
BASE COUNT 106 a 144 c 97 g 95 t
ORIGIN

Query Match 89.2%; Score 21.4; DB 9; Length 442;
Best Local Similarity 95.7%; Pred. No. 26;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || ||||||||||||||||||||
Db      23 GCGTTCTTTGCCAAGAGCGCCGC 45


RESULT 8
BM696193
LOCUS       BM696193                446 bp    mRNA    linear   EST 28-FEB-2002
DEFINITION  UI-E-CL1-afa-d-24-0-UI.r1 UI-E-CL1 Homo sapiens cDNA clone
            UI-E-CL1-afa-d-24-0-UI 5', mRNA sequence.
ACCESSION   BM696193
VERSION     BM696193.1  GI:19009451
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-E-CL1-afa-d-24-0-UI"
                     /clone_lib="UI-E-CL1"
                     /tissue_type="human retina"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-CL1 is a normalized cDNA library containing the
                     following tissue(s): retina. The library was constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tag for this library is CCGCG. This library was
                     created for the program, Gene Discovery in the Visual
                     System, supported by National Eye Institute (NEI)."
BASE COUNT       95 a    148 c    106 g     96 t      1 others
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 14;  Length 446;
  Best Local Similarity   95.7%;  Pred. No. 26;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || ||||||||||||||||||||
Db     226 GCGTTCTTTGCCAAGAGCGCCGC 248


RESULT 9
BM688227
LOCUS       BM688227                469 bp    mRNA    linear   EST 28-FEB-2002
DEFINITION  UI-E-CL0-aby-g-03-0-UI.r1 UI-E-CL0 Homo sapiens cDNA clone
            UI-E-CL0-aby-g-03-0-UI 5', mRNA sequence.
ACCESSION   BM688227
VERSION     BM688227.1  GI:19001478
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 469)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 425-459, >AT_rich#Low_complexity
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .469
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-E-CL0-aby-g-03-0-UI"
                     /clone_lib="UI-E-CL0"
                     /tissue_type="human retina"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-CL0 is a cDNA library containing the following
                     tissue(s): retina. The library was constructed according
                     to Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was ligated to an EcoR I adaptor, digested with Not
                     I, and cloned directionally into pT7T3-Pac vector. The
                     oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tag for this library is CCGCG. This library was
                     created for the program, Gene Discovery in the Visual
                     System, supported by National Eye Institute (NEI)."
BASE COUNT      110 a    155 c     99 g    104 t      1 others
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 14;  Length 469;
  Best Local Similarity   95.7%;  Pred. No. 27;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || ||||||||||||||||||||
Db      82 GCGTTCTTTGCCAAGAGCGCCGC 104


RESULT 10
AL712402
LOCUS       AL712402                471 bp    mRNA    linear   EST 22-MAR-2002
```

```
DEFINITION  DKFZp686N1788_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686N1788 5', mRNA sequence.
ACCESSION   AL712402
VERSION     AL712402.1  GI:19695757
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,W., Weil,B. and Wiemann,S.
  TITLE     EST (Wambutt,R., Heubner,D., Mewes,H.W., Weil,B. and Wiemann,S.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Wambutt R
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            No s1 sequence available.
            This clone (DKFZp686N1788) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686N1788"
                     /clone_lib="686 (synonym: hlcc3)"
                     /tissue_type="human skeletal muscle"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
BASE COUNT      109 a    158 c     99 g    105 t
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 9;  Length 471;
  Best Local Similarity   95.7%;  Pred. No. 27;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || ||||||||||||||||||||
Db      23 GCGTTCTTTGCCAAGAGCGCCGC 45


RESULT 11
BQ250368
LOCUS       BQ250368                 471 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE25006A11R TaE25 Triticum aestivum cDNA clone TaE25006A11R, mRNA
            sequence.
ACCESSION   BQ250368
VERSION     BQ250368.1  GI:20446244
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >870 bp
            Plate: 006  row: A  column: 11
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Triticum aestivum"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="TaE25006A11R"
                     /clone_lib="TaE25"
                     /tissue_type="developing seeds"
                     /dev_stage="25 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pCMV-SPORT6.0 (invitrogen Technologies);
                     Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
                     of cultivar Glenlea 25 days post-anthesis"
BASE COUNT      108 a    164 c    108 g     91 t
ORIGIN

  Query Match             89.2%;  Score 21.4;  DB 14;  Length 471;
  Best Local Similarity   95.7%;  Pred. No. 27;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCCAAGAGCGCCGC 23
           || ||||||||||||||||||||
Db     174 GCGTTCTTTGCCAAGAGCGCCGC 196


RESULT 12
BM688069
LOCUS       BM688069                 474 bp    mRNA    linear   EST 28-FEB-2002
DEFINITION  UI-E-CL0-abv-c-04-0-UI.r1 UI-E-CL0 Homo sapiens cDNA clone
            UI-E-CL0-abv-c-04-0-UI 5', mRNA sequence.
ACCESSION   BM688069
VERSION     BM688069.1  GI:19001327
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 474)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 335-367, >AT_rich#Low_complexity
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .474
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-E-CL0-abv-c-04-0-UI"
                     /clone_lib="UI-E-CL0"
                     /tissue_type="human retina"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
```

UI-E-CL0 is a cDNA library containing the following tissue(s): retina. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCGCG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 110 a 158 c 105 g 101 t
ORIGIN

Query Match 89.2%; Score 21.4; DB 14; Length 474;
Best Local Similarity 95.7%; Pred. No. 27;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCTTTCTTTGCCAAGAGCGCCGC 23
|| ||||||||||||||||||||||
Db 41 GCGTTCTTTGCCAAGAGCGCCGC 63

RESULT 13
BM662914/c
LOCUS BM662914 478 bp mRNA linear EST 27-FEB-2002
DEFINITION UI-E-CK0-aan-e-04-0-UI.s1 UI-E-CK0 Homo sapiens cDNA clone UI-E-CK0-aan-e-04-0-UI 3', mRNA sequence.
ACCESSION BM662914
VERSION BM662914.1 GI:18967816
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 478)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .478
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-CK0-aan-e-04-0-UI"
/clone_lib="UI-E-CK0"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-CK0 is a cDNA library containing the following tissue(s): Retina Foveal and Macular. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTCC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).
TAG_LIB=UI-E-CK0
TAG_TISSUE=Foveal and Macular Retina
TAG_SEQ=GTCC"

BASE COUNT 86 a 109 c 157 g 123 t 3 others
ORIGIN

Query Match 89.2%; Score 21.4; DB 13; Length 478;
Best Local Similarity 95.7%; Pred. No. 27;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCTTTCTTTGCCAAGAGCGCCGC 23
|| ||||||||||||||||||||||
Db 307 GCGTTCTTTGCCAAGAGCGCCGC 285

RESULT 14
BM690137
LOCUS BM690137 493 bp mRNA linear EST 28-FEB-2002
DEFINITION UI-E-CL0-acd-d-08-0-UI.r1 UI-E-CL0 Homo sapiens cDNA clone UI-E-CL0-acd-d-08-0-UI 5', mRNA sequence.
ACCESSION BM690137
VERSION BM690137.1 GI:19003395
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 493)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA sequence: 425-459, >AT_rich#Low_complexity
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .493
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-CL0-acd-d-08-0-UI"
/clone_lib="UI-E-CL0"
/tissue_type="human retina"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-CL0 is a cDNA library containing the following tissue(s): retina. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCGCG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 110 a 162 c 108 g 113 t
ORIGIN

Query Match 89.2%; Score 21.4; DB 14; Length 493;
Best Local Similarity 95.7%; Pred. No. 27;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCTTTCTTTGCCAAGAGCGCCGC 23
|| ||||||||||||||||||||
Db 131 GCGTTCTTTGCCAAGAGCGCCGC 153

RESULT 15
BM691592
LOCUS BM691592 493 bp mRNA linear EST 28-FEB-2002
DEFINITION UI-E-CI1-abh-c-07-0-UI.r1 UI-E-CI1 Homo sapiens cDNA clone UI-E-CI1-abh-c-07-0-UI 5', mRNA sequence.
ACCESSION BM691592
VERSION BM691592.1 GI:19004850
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 493)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA sequence: 302-335, >AT_rich#Low_complexity
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .493
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-CI1-abh-c-07-0-UI"
/clone_lib="UI-E-CI1"
/tissue_type="RPE and Choroid"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-CI1 is a normalized cDNA library containing the following tissue(s): RPE and Choroid. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 116 a 157 c 111 g 109 t
ORIGIN

Query Match 89.2%; Score 21.4; DB 14; Length 493;
Best Local Similarity 95.7%; Pred. No. 27;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCTTTCTTTGCCAAGAGCGCCGC 23
|| ||||||||||||||||||||
Db 8 GCGTTCTTTGCCAAGAGCGCCGC 30

Search completed: March 17, 2003, 13:09:15
Job time : 851.387 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:43:42 ; Search time 202.495 Seconds
(without alignments)
3161.870 Million cell updates/sec

Title: US-09-836-439-4
Perfect score: 22
Sequence: 1 aagaaaaaatctagacaagcaa 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		20.4	92.7	133095	2	AC046146	AC046146 Mus muscu
2		20.4	92.7	152147	2	AC095531	AC095531 Rattus no
3	c	20.4	92.7	172497	2	AC117913	AC117913 Rattus no
4		20.4	92.7	188795	2	AC113320	AC113320 Mus muscu
5	c	19.4	88.2	179767	9	AC069223	AC069223 Homo sapi
6		19.4	88.2	180638	9	AC068763	AC068763 Homo sapi
7		19.4	88.2	185841	2	AC027080	AC027080 Homo sapi
8	c	19.4	88.2	201416	2	AC130437	AC130437 Homo sapi
9		19.4	88.2	218859	2	AC022912	AC022912 Homo sapi
10	c	19.4	88.2	274349	2	AC093623	AC093623 Homo sapi
11	c	19	86.4	113940	2	AP005572	AP005572 Oryza sat
12	c	18.8	85.5	2052	5	AF012746	AF012746 Danio rer
13	c	18.8	85.5	2209	8	SCYPL252C	Z73608 S.cerevisia
14	c	18.8	85.5	2411	8	SCYPL253C	Z73609 S.cerevisia
15	c	18.8	85.5	37808	8	SC38KCXVI	Z67751 S.cerevisia
16		18.8	85.5	41578	3	CBRG46G14	AC091216 Caenorhab
17		18.8	85.5	52495	2	AC109155	AC109155 Mus muscu
18	c	18.8	85.5	52912	2	AC130512	AC130512 Rattus no
19	c	18.8	85.5	63484	2	AC108916	AC108916 Mus muscu
20	c	18.8	85.5	64417	2	AC131268	AC131268 Homo sapi
21		18.8	85.5	96036	2	AC119622	AC119622 Rattus no
22		18.8	85.5	97454	9	AL139147	AL139147 Human DNA
23		18.8	85.5	99374	9	AC092031	AC092031 Homo sapi
24	c	18.8	85.5	112632	2	AL137158	AL137158 Homo sapi
25		18.8	85.5	113955	9	AC068291	AC068291 Homo sapi
26		18.8	85.5	114457	2	AC095684	AC095684 Rattus no
27		18.8	85.5	120527	2	AP005487	AP005487 Oryza sat
28	c	18.8	85.5	132072	9	AC004841	AC004841 Homo sapi
29		18.8	85.5	140658	5	AF112374	AF112374 Danio rer
30		18.8	85.5	141138	2	AL845428	AL845428 Danio rer
31	c	18.8	85.5	143496	2	AC111035	AC111035 Mus muscu
32	c	18.8	85.5	143823	9	AC019195	AC019195 Homo sapi
33	c	18.8	85.5	151040	2	AC068389	AC068389 Homo sapi
34		18.8	85.5	152161	9	AL359074	AL359074 Human DNA
35		18.8	85.5	154348	2	AC069480	AC069480 Homo sapi
36	c	18.8	85.5	156195	9	AC093799	AC093799 Homo sapi
37		18.8	85.5	157038	2	AC099275	AC099275 Rattus no
38		18.8	85.5	157823	2	AL161795	AL161795 Homo sapi
39		18.8	85.5	158075	2	AC123076	AC123076 Rattus no
40		18.8	85.5	159386	2	AC098370	AC098370 Rattus no
41		18.8	85.5	163830	2	AC068235	AC068235 Homo sapi
42		18.8	85.5	166292	2	CNS08C8S	AL731888 Oryza sat
43		18.8	85.5	166764	2	AC115875	AC115875 Mus muscu
44	c	18.8	85.5	168199	9	CNS01DTS	AL132857 Human chr
45		18.8	85.5	169295	9	AL732602	AL732602 Human DNA

ALIGNMENTS

RESULT 1
AC046146
LOCUS AC046146 133095 bp DNA linear HTG 16-OCT-2001
DEFINITION Mus musculus chromosome 12 clone RP23-321N21, *** SEQUENCING IN PROGRESS ***, 31 unordered pieces.
ACCESSION AC046146
VERSION AC046146.6 GI:16118085
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 133095)
AUTHORS Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C., Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,

```
            Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
            Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
            Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
            Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
            Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
            Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
            Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
            Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
            Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
            Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
            Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
            Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
            Worley,K. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 133095)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 14, 2001 this sequence version replaced gi:11094634.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: MADV
                Center clone name: RP23-321N21
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 94% of reads
                Chemistry: Dye-terminator Big Dye: 6% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 118378 bases at least Q40
                Consensus quality: 144876 bases at least Q30
                Consensus quality: 155160 bases at least Q20
                Estimated insert size: 148882; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 31 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    10336: contig of 10336 bp in length
            *    10337    10436: gap of unknown length
            *    10437    19208: contig of 8772 bp in length
            *    19209    19308: gap of unknown length
            *    19309    26724: contig of 7416 bp in length
            *    26725    26824: gap of unknown length
            *    26825    33512: contig of 6688 bp in length
            *    33513    33612: gap of unknown length
            *    33613    39388: contig of 5776 bp in length
            *    39389    39488: gap of unknown length
            *    39489    45330: contig of 5842 bp in length
            *    45331    45430: gap of unknown length
            *    45431    50898: contig of 5468 bp in length
            *    50899    50998: gap of unknown length
            *    50999    56466: contig of 5468 bp in length
            *    56467    56566: gap of unknown length
            *    56567    62096: contig of 5530 bp in length
            *    62097    62196: gap of unknown length
            *    62197    65971: contig of 3775 bp in length
            *    65972    66071: gap of unknown length
            *    66072    70109: contig of 4038 bp in length
            *    70110    70209: gap of unknown length
            *    70210    72595: contig of 2386 bp in length
            *    72596    72695: gap of unknown length
            *    72696    74972: contig of 2277 bp in length
            *    74973    75072: gap of unknown length
            *    75073    79255: contig of 4183 bp in length
            *    79256    79355: gap of unknown length
            *    79356    83183: contig of 3828 bp in length
            *    83184    83283: gap of unknown length
            *    83284    86781: contig of 3498 bp in length
            *    86782    86881: gap of unknown length
            *    86882    90021: contig of 3140 bp in length
            *    90022    90121: gap of unknown length
            *    90122    93597: contig of 3476 bp in length
            *    93598    93697: gap of unknown length
            *    93698    97229: contig of 3532 bp in length
            *    97230    97329: gap of unknown length
            *    97330   100516: contig of 3187 bp in length
            *   100517   100616: gap of unknown length
            *   100617   104190: contig of 3574 bp in length
            *   104191   104290: gap of unknown length
            *   104291   107955: contig of 3665 bp in length
            *   107956   108055: gap of unknown length
            *   108056   111930: contig of 3875 bp in length
            *   111931   112030: gap of unknown length
            *   112031   114125: contig of 2095 bp in length
            *   114126   114225: gap of unknown length
            *   114226   116520: contig of 2295 bp in length
            *   116521   116620: gap of unknown length
            *   116621   118829: contig of 2209 bp in length
            *   118830   118929: gap of unknown length
            *   118930   121281: contig of 2352 bp in length
            *   121282   121381: gap of unknown length
            *   121382   123762: contig of 2381 bp in length
            *   123763   123862: gap of unknown length
            *   123863   126536: contig of 2674 bp in length
            *   126537   126636: gap of unknown length
            *   126637   130483: contig of 3847 bp in length
            *   130484   130583: gap of unknown length
            *   130584   133095: contig of 2512 bp in length.
FEATURES             Location/Qualifiers
     source          1. .133095
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="12"
                     /clone="RP23-321N21"
BASE COUNT    34946 a  30809 c  29300 g  35021 t   3019 others
ORIGIN

  Query Match             92.7%;  Score 20.4;  DB 2;  Length 133095;
  Best Local Similarity   95.5%;  Pred. No. 65;
  Matches   21;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           |||| ||||||||||||||||||
Db   71615 AAGACAAAATCTAGACAAGCAA 71636


RESULT 2
AC095531
LOCUS       AC095531              152147 bp    DNA     linear   HTG 10-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-8N2, *** SEQUENCING IN PROGRESS ***,
            58 unordered pieces.
ACCESSION   AC095531
VERSION     AC095531.3  GI:21716992
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 152147)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
```

Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 152147)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 152147)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 9, 2002 this sequence version replaced gi:17942049.

```
-------------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
    Center project name: GCPY
    Center clone name: CH230-8N2
-------------- Summary Statistics
    Sequencing vector: Plasmid;
    Chemistry: Dye-terminator Big Dye: 100% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 92864 bases at least Q40
    Consensus quality: 97631 bases at least Q30
    Consensus quality: 101839 bases at least Q20
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 58 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1     1151: contig of 1151 bp in length
*    1152     1251: gap of unknown length
*    1252     2705: contig of 1454 bp in length
*    2706     2805: gap of unknown length
*    2806     4321: contig of 1516 bp in length
*    4322     4421: gap of unknown length
*    4422     5859: contig of 1438 bp in length
*    5860     5959: gap of unknown length
*    5960     7226: contig of 1267 bp in length
*    7227     7326: gap of unknown length
*    7327     8443: contig of 1117 bp in length
*    8444     8543: gap of unknown length
*    8544     9703: contig of 1160 bp in length
*    9704     9803: gap of unknown length
*    9804    11035: contig of 1232 bp in length
*   11036    11135: gap of unknown length
*   11136    12923: contig of 1788 bp in length
*   12924    13023: gap of unknown length
*   13024    14175: contig of 1152 bp in length
*   14176    14275: gap of unknown length
*   14276    16212: contig of 1937 bp in length
*   16213    16312: gap of unknown length
*   16313    17891: contig of 1579 bp in length
*   17892    17991: gap of unknown length
*   17992    19332: contig of 1341 bp in length
*   19333    19432: gap of unknown length
*   19433    20845: contig of 1413 bp in length
*   20846    20945: gap of unknown length
*   20946    22783: contig of 1838 bp in length
*   22784    22883: gap of unknown length
*   22884    23919: contig of 1036 bp in length
*   23920    24019: gap of unknown length
*   24020    25188: contig of 1169 bp in length
*   25189    25288: gap of unknown length
*   25289    26434: contig of 1146 bp in length
*   26435    26534: gap of unknown length
*   26535    27939: contig of 1405 bp in length
*   27940    28039: gap of unknown length
*   28040    29820: contig of 1781 bp in length
*   29821    29920: gap of unknown length
*   29921    31409: contig of 1489 bp in length
*   31410    31509: gap of unknown length
*   31510    33307: contig of 1798 bp in length
*   33308    33407: gap of unknown length
*   33408    35426: contig of 2019 bp in length
*   35427    35526: gap of unknown length
*   35527    36880: contig of 1354 bp in length
*   36881    36980: gap of unknown length
*   36981    38529: contig of 1549 bp in length
*   38530    38629: gap of unknown length
*   38630    40146: contig of 1517 bp in length
*   40147    40246: gap of unknown length
*   40247    42517: contig of 2271 bp in length
*   42518    42617: gap of unknown length
*   42618    44292: contig of 1675 bp in length
*   44293    44392: gap of unknown length
*   44393    46708: contig of 2316 bp in length
*   46709    46808: gap of unknown length
*   46809    49242: contig of 2434 bp in length
*   49243    49342: gap of unknown length
*   49343    51258: contig of 1916 bp in length
*   51259    51358: gap of unknown length
*   51359    53360: contig of 2002 bp in length
*   53361    53460: gap of unknown length
*   53461    55148: contig of 1688 bp in length
*   55149    55248: gap of unknown length
```

```
            *    55249    58677: contig of 3429 bp in length
            *    58678    58777: gap of unknown length
            *    58778    60637: contig of 1860 bp in length
            *    60638    60737: gap of unknown length
            *    60738    63930: contig of 3193 bp in length
            *    63931    64030: gap of unknown length
            *    64031    66398: contig of 2368 bp in length
            *    66399    66498: gap of unknown length
            *    66499    69978: contig of 3480 bp in length
            *    69979    70078: gap of unknown length
            *    70079    72552: contig of 2474 bp in length
            *    72553    72652: gap of unknown length
            *    72653    74752: contig of 2100 bp in length
            *    74753    74852: gap of unknown length
            *    74853    77949: contig of 3097 bp in length
            *    77950    78049: gap of unknown length
            *    78050    79999: contig of 1950 bp in length
            *    80000    80099: gap of unknown length
            *    80100    82883: contig of 2784 bp in length
            *    82884    82983: gap of unknown length
            *    82984    86240: contig of 3257 bp in length
            *    86241    86340: gap of unknown length
            *    86341    90145: contig of 3805 bp in length
            *    90146    90245: gap of unknown length
            *    90246    92780: contig of 2535 bp in length
            *    92781    92880: gap of unknown length
            *    92881    97022: contig of 4142 bp in length
            *    97023    97122: gap of unknown length
            *    97123    99914: contig of 2792 bp in length
            *    99915   100014: gap of unknown length
            *   100015   103192: contig of 3178 bp in length
            *   103193   103292: gap of unknown length
            *   103293   107022: contig of 3730 bp in length
            *   107023   107122: gap of unknown length
            *   107123   110597: contig of 3475 bp in length
            *   110598   110697: gap of unknown length
            *   110698   116077: contig of 5380 bp in length
            *   116078   116177: gap of unknown length
            *   116178   120578: contig of 4401 bp in length
            *   120579   120678: gap of unknown length
            *   120679   125200: contig of 4522 bp in length

  Query Match             92.7%;  Score 20.4;  DB 2;  Length 152147;
  Best Local Similarity   95.5%;  Pred. No. 63;
  Matches   21;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           |||||||||| ||||||||||
Db 141943 AAGAAAAAATGTAGACAAGCAA 141964


RESULT 3
AC117913/c
LOCUS       AC117913              172497 bp    DNA     linear   HTG 18-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-35D24, *** SEQUENCING IN PROGRESS
            ***, 59 unordered pieces.
ACCESSION   AC117913
VERSION     AC117913.4  GI:21747159
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 172497)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172497)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 172497)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 14, 2002 this sequence version replaced gi:20258282.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GVPM
                Center clone name: CH230-35D24
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 118901 bases at least Q40
                Consensus quality: 125278 bases at least Q30
                Consensus quality: 129335 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 59 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
*       1    1133: contig of 1133 bp in length
*    1134    1233: gap of unknown length
*    1234    2457: contig of 1224 bp in length
*    2458    2557: gap of unknown length
*    2558    3893: contig of 1336 bp in length
*    3894    3993: gap of unknown length
*    3994    5079: contig of 1086 bp in length
*    5080    5179: gap of unknown length
*    5180    6739: contig of 1560 bp in length
*    6740    6839: gap of unknown length
*    6840    8037: contig of 1198 bp in length
*    8038    8137: gap of unknown length
*    8138    9769: contig of 1632 bp in length
*    9770    9869: gap of unknown length
*    9870   11173: contig of 1304 bp in length
*   11174   11273: gap of unknown length
*   11274   12329: contig of 1056 bp in length
*   12330   12429: gap of unknown length
*   12430   14363: contig of 1934 bp in length
*   14364   14463: gap of unknown length
*   14464   15857: contig of 1394 bp in length
*   15858   15957: gap of unknown length
*   15958   17164: contig of 1207 bp in length
*   17165   17264: gap of unknown length
*   17265   18407: contig of 1143 bp in length
*   18408   18507: gap of unknown length
*   18508   20371: contig of 1864 bp in length
*   20372   20471: gap of unknown length
*   20472   21792: contig of 1321 bp in length
*   21793   21892: gap of unknown length
*   21893   23539: contig of 1647 bp in length
*   23540   23639: gap of unknown length
*   23640   24789: contig of 1150 bp in length
*   24790   24889: gap of unknown length
*   24890   26903: contig of 2014 bp in length
*   26904   27003: gap of unknown length
*   27004   28488: contig of 1485 bp in length
*   28489   28588: gap of unknown length
*   28589   29829: contig of 1241 bp in length
*   29830   29929: gap of unknown length
*   29930   31462: contig of 1533 bp in length
*   31463   31562: gap of unknown length
*   31563   33655: contig of 2093 bp in length
*   33656   33755: gap of unknown length
*   33756   35134: contig of 1379 bp in length
*   35135   35234: gap of unknown length
*   35235   37277: contig of 2043 bp in length
*   37278   37377: gap of unknown length
*   37378   38782: contig of 1405 bp in length
*   38783   38882: gap of unknown length
*   38883   40412: contig of 1530 bp in length
*   40413   40512: gap of unknown length
*   40513   42065: contig of 1553 bp in length
*   42066   42165: gap of unknown length
*   42166   44423: contig of 2258 bp in length
*   44424   44523: gap of unknown length
*   44524   45924: contig of 1401 bp in length
*   45925   46024: gap of unknown length
*   46025   47783: contig of 1759 bp in length
*   47784   47883: gap of unknown length
*   47884   50262: contig of 2379 bp in length
*   50263   50362: gap of unknown length
*   50363   52442: contig of 2080 bp in length
*   52443   52542: gap of unknown length
*   52543   55030: contig of 2488 bp in length
*   55031   55130: gap of unknown length
*   55131   57583: contig of 2453 bp in length
*   57584   57683: gap of unknown length
*   57684   59735: contig of 2052 bp in length
*   59736   59835: gap of unknown length
*   59836   62023: contig of 2188 bp in length
*   62024   62123: gap of unknown length
*   62124   65327: contig of 3204 bp in length
*   65328   65427: gap of unknown length
*   65428   67771: contig of 2344 bp in length
*   67772   67871: gap of unknown length
*   67872   70986: contig of 3115 bp in length
*   70987   71086: gap of unknown length
*   71087   74496: contig of 3410 bp in length
*   74497   74596: gap of unknown length
*   74597   76391: contig of 1795 bp in length
*   76392   76491: gap of unknown length
*   76492   80570: contig of 4079 bp in length
*   80571   80670: gap of unknown length
*   80671   84667: contig of 3997 bp in length
*   84668   84767: gap of unknown length
*   84768   88261: contig of 3494 bp in length
*   88262   88361: gap of unknown length
*   88362   91664: contig of 3303 bp in length
*   91665   91764: gap of unknown length
*   91765   96059: contig of 4295 bp in length
*   96060   96159: gap of unknown length
*   96160  101392: contig of 5233 bp in length
*  101393  101492: gap of unknown length
*  101493  104583: contig of 3091 bp in length
*  104584  104683: gap of unknown length
*  104684  109409: contig of 4726 bp in length
*  109410  109509: gap of unknown length
*  109510  114694: contig of 5185 bp in length
*  114695  114794: gap of unknown length
*  114795  119788: contig of 4994 bp in length
*  119789  119888: gap of unknown length
*  119889  126088: contig of 6200 bp in length
*  126089  126188: gap of unknown length
*  126189  134246: contig of 8058 bp in length
*  134247  134346: gap of unknown length
*  134347  140261: contig of 5915 bp in length

  Query Match             92.7%;  Score 20.4;  DB 2;  Length 172497;
  Best Local Similarity   95.5%;  Pred. No. 62;
  Matches   21;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           |||||||||| ||||||||||
Db   45058 AAGAAAAAATGTAGACAAGCAA 45037


RESULT 4
AC113320
LOCUS       AC113320              188795 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Mus musculus clone RP23-445E16, WORKING DRAFT SEQUENCE, 11 ordered
            pieces.
ACCESSION   AC113320
VERSION     AC113320.2  GI:22474991
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 188795)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-445E16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 188795)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
```

```
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre;N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 188795)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 24, 2002 this sequence version replaced gi:18997726.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22784
                Center clone name: 445_E_16
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 185600 bases at least Q40
                Consensus quality: 186954 bases at least Q30
                Consensus quality: 187365 bases at least Q20
                Insert size: 188000; agarose-fp
                Insert size: 187795; sum-of-contigs
                Quality coverage: 9.1 in Q20 bases; agarose-fp
                Quality coverage: 9.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    47353: contig of 47353 bp in length
            *    47354 47453: gap of      100 bp
            *    47454    48767: contig of 1314 bp in length
            *    48768 48867: gap of      100 bp
            *    48868    50642: contig of 1775 bp in length
            *    50643 50742: gap of      100 bp
            *    50743    51985: contig of 1243 bp in length
            *    51986 52085: gap of      100 bp
            *    52086    54528: contig of 2443 bp in length
            *    54529 54628: gap of      100 bp
            *    54629    58116: contig of 3488 bp in length
            *    58117 58216: gap of      100 bp
            *    58217    60812: contig of 2596 bp in length
            *    60813 60912: gap of      100 bp
            *    60913    62765: contig of 1853 bp in length
            *    62766 62865: gap of      100 bp
            *    62866    72529: contig of 9664 bp in length
            *    72530 72629: gap of      100 bp
            *    72630    96147: contig of 23518 bp in length
            *    96148 96247: gap of      100 bp
            *    96248   188795: contig of 92548 bp in length.
FEATURES             Location/Qualifiers
     source          1. .188795
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-445E16"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .47353
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    47454. .48767
                     /note="assembly_fragment"
     misc_feature    48868. .50642
                     /note="assembly_fragment"
     misc_feature    50743. .51985
                     /note="assembly_fragment"
     misc_feature    52086. .54528
                     /note="assembly_fragment"
     misc_feature    54629. .58116
                     /note="assembly_fragment"
     misc_feature    58217. .60812
                     /note="assembly_fragment"
     misc_feature    60913. .62765
                     /note="assembly_fragment"
     misc_feature    62866. .72529
                     /note="assembly_fragment"
     misc_feature    72630. .96147
                     /note="assembly_fragment"
     misc_feature    96248. .188795
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    57294 a  37412 c  36369 g  56709 t   1011 others
ORIGIN

  Query Match             92.7%;  Score 20.4;  DB 2;  Length 188795;
  Best Local Similarity   95.5%;  Pred. No. 61;
  Matches   21;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           |||||||||||||||||| |||
Db   36223 AAGAAAAAATCTAGACAAACAA 36244


RESULT 5
AC069223/c
LOCUS       AC069223              179767 bp    DNA     linear   PRI 30-MAR-2001
DEFINITION  Homo sapiens 3 BAC RP11-398O21 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC069223
VERSION     AC069223.15  GI:13489121
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179767)
```

AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L. and Gibbs,R.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 179767)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (22-MAY-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 179767)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (30-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Mar 30, 2001 this sequence version replaced gi:13430920.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT----------

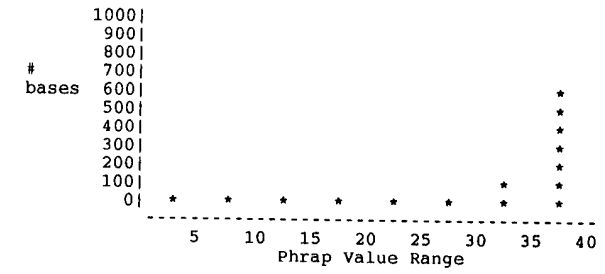
--- Summary Statistics ---

Contig length:	179767
Phrap values in estimate:	179551
Average error rate (BCM-Phrap estimate):	4.4907e-06
Fraction of Phrap values less than 40 :	0.00499023
Number of consensus changing edits:	4
Number of N's in consensus :	0

--- Consensus changing edits ---

Position	Original+Context	Edited+Context
3	atccaaanaa(n)tctgatcttc	atccgcggaa(t)tctgatcttc
107934	agcctctttt(n)nacagagact	agcctctttt(t)tacagagact
107935	gcctcttttn(n)acagagacta	gcctcttttt(t)acagagacta
159144	atggcaatca(n)cacacatggt	atggcaatca(c)cacacatggt

--- Distribution of Quality < 40 Bases ---



Version: 1.01 qxfo.

FEATURES	Location/Qualifiers
source	1. .179767
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="3"
	/clone="RP11-398O21"
repeat_region	55. .358
	/rpt_family="AluY"
repeat_region	complement(639. .774)
	/rpt_family="MIR"
repeat_region	1202. .1410
	/rpt_family="AluJb"
repeat_region	complement(4054. .4428)
	/rpt_family="L2"
repeat_region	complement(4951. .5263)

```
                     /rpt_family="AluSq"
     repeat_region   7302. .7547
                     /rpt_family="L1PA13"
     repeat_region   complement(10047. .10274)
                     /rpt_family="MIR"
     repeat_region   complement(10755. .10914)
                     /rpt_family="HAL1"
     repeat_region   complement(10947. .11385)
                     /rpt_family="HAL1"
     repeat_region   complement(12041. .12199)
                     /rpt_family="AluSp"
     repeat_region   12244. .12383
                     /rpt_family="MIR"
     repeat_region   complement(13142. .13309)
                     /rpt_family="MER5B"
     repeat_region   14874. .15137
                     /rpt_family="AluJo"
     repeat_region   complement(16376. .16594)
                     /rpt_family="L2"
     misc_feature    17830. .19096
                     /note="Region similar to Homo sapiens mRNA; cDNA
                     DKFZp586L2024 (from clone DKFZp586L2024); partial cds
                     /cds=(0,1122) /gb=AL117664 /gi=5912260 /ug=Hs.58419
                     /len=2298"
     STS             18389. .18517
                     /standard_name="G21620"
                     /db_xref="dbSTS:35458"
     misc_feature    complement(19478. .20298)
                     /note="Region similar to Homo sapiens TRK-fused gene
                     (TFG), mRNA /cds=(18,1220) /gb=NM_006070 /gi=5174718
                     /ug=Hs.250897 /len=1677"
     STS             19596. .19757
                     /standard_name="NIB1880"
     repeat_region   22027. .22107
                     /rpt_family="MIR"
     repeat_region   22097. .22217
                     /rpt_family="MER5B"
     repeat_region   complement(22479. .22694)
                     /rpt_family="MER58A"
     repeat_region   23342. .23375
                     /rpt_family="AT_rich"
     repeat_region   complement(25239. .25373)
                     /rpt_family="MIR"

  Query Match             88.2%;  Score 19.4;  DB 9;  Length 179767;
  Best Local Similarity   95.2%;  Pred. No. 1.6e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCA 21
           ||||||||||||||||| |||
Db    4959 AAGAAAAAATCTAGACAGGCA 4939


RESULT 6
AC068763
LOCUS       AC068763               180638 bp    DNA     linear   PRI 25-JUL-2002
DEFINITION  Homo sapiens 3 BAC RP11-133K20 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC068763
VERSION     AC068763.11  GI:14669935
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180638)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
            Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S.,
            Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180638)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 180638)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 180638)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 180638)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 11, 2001 this sequence version replaced gi:13493008.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
```

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT----------

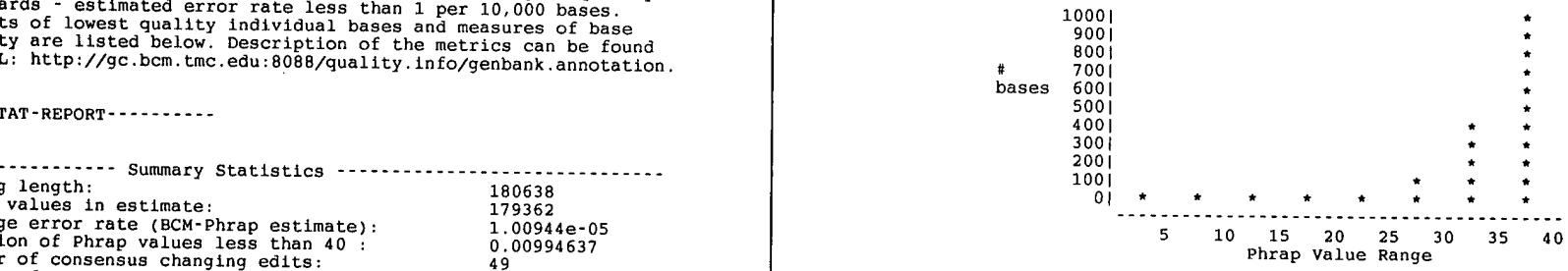
---------- Summary Statistics ----------

Contig length:	180638
Phrap values in estimate:	179362
Average error rate (BCM-Phrap estimate):	1.00944e-05
Fraction of Phrap values less than 40 :	0.00994637
Number of consensus changing edits:	49
Number of N's in consensus :	0

---------- Consensus changing edits ----------

Position	Original+Context	Edited+Context
17175	tgtttataga(n)aaaagagtcc	tgtttataga(a)aaaagagtcc
17467	tatgagtgag(n)acgtgcggtg	tatgagtgag(a)acgtgcggtg
19821	atctagaaga(n)ntggataaat	atctagaaga(a)atggataaat
19822	tctagaagan(n)tggataaatt	tctagaagaa(a)tggataaatt
29835	gatggcttta(n)gatgtgttta	gatggcttta(g)gatgtgttta
49503	acttcaggta(n)actgctcagc	acttcaggta(a)actgctcagc
51340	tctcattatt(n)gtatttctac	tctcattatt(t)gtatttctac
52598	ggaaaataat(n)catgtacaaa	ggaaaataat(a)catgtacaaa
60642	acttaaaaat(n)aaattttcgg	acttaaaaat(t)aaattttcgg
60920	caacaagagc(t)tnactccatc	caacaagagc(g)aaactccatc
60921	aacaagagct(t)nactccatct	aacaagagcg(a)aactccatct
60922	acaagagctt(n)actccatctc	acaagagcga(a)actccatctc
69475	agatcgcgcc(a)ccgcactcca	agatcgcgcc(c)ccgcactcca
73510	atctcagctc(n)ctgcaacttc	atctcagctc(a)ctgcaacttc
73543	tcaggagatt(n)tcctgcctca	tcaggagatt(c)tcctgcctca
73554	tcctgcctca(n)cctcccaaat	tcctgcctca(g)cctcccaaat
73566	ctcccaaata(n)ntgggattac	ctcccaaata(g)ctgggattac
73567	tcccaaatan(n)tgggattaca	tcccaaatag(c)tgggattaca
73596	ccaccatgcc(n)ggctaatttt	ccaccatgcc(t)ggctaatttt
73704	gtgttgggat(n)acaggcgtga	gtgttgggat(t)acaggcgtga
73721	gtgagccacc(n)ngcctggcct	gtgagccacc(g)tgcctggcct
73722	tgagccaccn(n)gcctggcctc	tgagccaccg(t)gcctggcctc
73748	ttctttatta(n)tntgagtagt	ttctttatta(t)tctgagtagt
73750	ctttattant(n)tgagtagtgg	ctttattatt(c)tgagtagtgg
73796	accagctcct(n)gattcattga	accagctcct(g)gattcattga
101317	ctgtaagctt(n)ttaacactca	ctgtaagctt(t)ttaacactca
104897	ttttttttt(g)gagagagtct	ttttttttt(t)gagagagtct
109869	gtgggaggga(n)gangggtgg	gtgggaggga(g)gaggggtgg
109872	ggaggganga(n)ggggtgggag	ggagggagga(g)ggggtgggag
113503	gttgcctcca(n)nccattttaa	gttgcctcca(g)accattttaa
113504	ttgcctccan(n)ccattttaac	ttgcctccag(a)ccattttaac
125301	aaatacctat(n)ataattatat	aaatacctat(t)ataattatat
128519	ctgataattc(n)gctttttttc	ctgataattc(t)gctttttttc
134437	ttccgggatt(n)tcctcctaaa	ttccgggatt(t)tcctcctaaa
134462	cagctgcttg(n)ctagccataa	cagctgcttg(t)ctagccataa
137769	cctggctagg(n)gtggngcctc	cctggctagg(t)gtggtgcctc
137774	ctaggngtgg(n)gcctcacgcc	ctaggtgtgg(t)gcctcacgcc
141481	tagcacaatg(n)actttaagct	tagcacaatg(t)actttaagct
151926	atactcagaa(n)aaaatccaaa	atactcagaa(t)aaaatccaaa
180507	tcaccaggtg(n)gnnnnnnnnt	tcaccaggtg(g)gatttcattt
180509	accaggtgng(n)nnnnnnntat	accaggtggg(a)tttcatttat
180510	ccaggtgngn(n)nnnnnntatg	ccaggtggga(t)ttcatttatg
180511	caggtgngnn(n)nnnnntatgg	caggtgggat(t)tcatttatgg
180512	aggtgngnnn(n)nnnntatggt	aggtgggatt(t)catttatggt
180513	ggtgngnnnn(n)nnntatggta	ggtgggattt(c)atttatggta
180514	gtgngnnnnn(n)nntatggtaa	gtgggatttc(a)tttatggtaa
180515	tgngnnnnnn(n)ntatggtaac	tgggatttca(t)ttatggtaac
180516	gngnnnnnnn(n)tatggtaact	gggatttcat(t)tatggtaact
180569	ttgtaattca(n)aaaaaatttt	ttgtaattca(a)aaaaaatttt

--------- Distribution of Quality < 40 Bases ----------



Version: 1.01 qxfo.

FEATURES Location/Qualifiers

Query Match 88.2%; Score 19.4; DB 9; Length 180638;
Best Local Similarity 95.2%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 AAGAAAAAATCTAGACAAGCA 21
          |||||||||||||||||  |||
Db 163480 AAGAAAAAATCTAGACAGGCA 163500
```

RESULT 7
AC027080
LOCUS AC027080 185841 bp DNA linear HTG 19-APR-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-548O23 map 3, WORKING DRAFT SEQUENCE, 22 unordered pieces.
ACCESSION AC027080
VERSION AC027080.2 GI:7596901
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 185841)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 3, clone RP11-548O23
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 185841)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,

```
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 19, 2000 this sequence version replaced gi:7329443.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8201
                Center clone name: 548_O_23
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 172210 bases at least Q40
                Consensus quality: 178990 bases at least Q30
                Consensus quality: 181814 bases at least Q20
                Insert size: 189000; agarose-fp
                Insert size: 183741; sum-of-contigs
                Quality coverage: 4.0 in Q20 bases; agarose-fp
                Quality coverage: 4.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1458: contig of 1458 bp in length
            *     1459 1558: gap of      100 bp
            *     1559     3252: contig of 1694 bp in length
            *     3253 3352: gap of      100 bp
            *     3353     5248: contig of 1896 bp in length
            *     5249 5348: gap of      100 bp
            *     5349     8165: contig of 2817 bp in length
            *     8166 8265: gap of      100 bp
            *     8266    11611: contig of 3346 bp in length
            *    11612 11711: gap of      100 bp
            *    11712    16128: contig of 4417 bp in length
            *    16129 16228: gap of      100 bp
            *    16229    22113: contig of 5885 bp in length
            *    22114 22213: gap of      100 bp
            *    22214    28574: contig of 6361 bp in length
            *    28575 28674: gap of      100 bp
            *    28675    34052: contig of 5378 bp in length
            *    34053 34152: gap of      100 bp
            *    34153    41388: contig of 7236 bp in length
            *    41389 41488: gap of      100 bp
            *    41489    50939: contig of 9451 bp in length
            *    50940 51039: gap of      100 bp
            *    51040    61135: contig of 10096 bp in length
            *    61136 61235: gap of      100 bp
            *    61236    69782: contig of 8547 bp in length
            *    69783 69882: gap of      100 bp
            *    69883    77230: contig of 7348 bp in length
            *    77231 77330: gap of      100 bp
            *    77331    86536: contig of 9206 bp in length
            *    86537 86636: gap of      100 bp
            *    86637    98162: contig of 11526 bp in length
            *    98163 98262: gap of      100 bp
            *    98263   109460: contig of 11198 bp in length
            *   109461 109560: gap of      100 bp
            *   109561   121285: contig of 11725 bp in length
            *   121286 121385: gap of      100 bp
            *   121386   135542: contig of 14157 bp in length
            *   135543 135642: gap of      100 bp
            *   135643   147139: contig of 11497 bp in length
            *   147140 147239: gap of      100 bp
            *   147240   161924: contig of 14685 bp in length
            *   161925 162024: gap of      100 bp
            *   162025   185841: contig of 23817 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185841
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-548O23"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1458
                     /note="assembly_fragment"
     misc_feature    1559. .3252
                     /note="assembly_fragment"
     misc_feature    3353. .5248
                     /note="assembly_fragment"
     misc_feature    5349. .8165
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    8266. .11611
                     /note="assembly_fragment"
     misc_feature    11712. .16128
                     /note="assembly_fragment"
     misc_feature    16229. .22113
                     /note="assembly_fragment"
     misc_feature    22214. .28574
                     /note="assembly_fragment"
     misc_feature    28675. .34052
                     /note="assembly_fragment"
     misc_feature    34153. .41388
                     /note="assembly_fragment"
     misc_feature    41489. .50939
                     /note="assembly_fragment"
     misc_feature    51040. .61135
                     /note="assembly_fragment"
     misc_feature    61236. .69782
                     /note="assembly_fragment"
     misc_feature    69883. .77230
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    77331. .86536
                     /note="assembly_fragment"
     misc_feature    86637. .98162
                     /note="assembly_fragment"
     misc_feature    98263. .109460
                     /note="assembly_fragment"
     misc_feature    109561. .121285
                     /note="assembly_fragment"
     misc_feature    121386. .135542
                     /note="assembly_fragment"
     misc_feature    135643. .147139
                     /note="assembly_fragment"
     misc_feature    147240. .161924
                     /note="assembly_fragment"
     misc_feature    162025. .185841
                     /note="assembly_fragment"
BASE COUNT    57091 a  34837 c  33673 g  58130 t   2110 others
```

```
ORIGIN

  Query Match             88.2%;  Score 19.4;  DB 2;  Length 185841;
  Best Local Similarity   95.2%;  Pred. No. 1.6e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 AAGAAAAAATCTAGACAAGCA 21
          ||||||||||||||||| |||
Db 127736 AAGAAAAAATCTAGACAGGCA 127756


RESULT 8
AC130437/c
LOCUS       AC130437              201416 bp    DNA     linear   HTG 10-AUG-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-740L19, *** SEQUENCING IN
            PROGRESS ***, 10 unordered pieces.
ACCESSION   AC130437
VERSION     AC130437.1  GI:22203194
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 201416)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 201416)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     ---------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
            Center project name: HDIU
            Center clone name: RP11-740L19
            ---------- Summary Statistics
            Chemistry: Dye-primer Bodipy: Infinity% of reads
            Chemistry: Dye-terminator Big Dye: Infinity% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 200930 bases at least Q40
            Consensus quality: 216627 bases at least Q30
            Consensus quality: 233347 bases at least Q20
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2055: contig of 2055 bp in length
            *     2056     2155: gap of unknown length
            *     2156     4275: contig of 2120 bp in length
            *     4276     4375: gap of unknown length
            *     4376     9623: contig of 5248 bp in length
            *     9624     9723: gap of unknown length
            *     9724    26423: contig of 16700 bp in length
            *    26424    26523: gap of unknown length
            *    26524    42616: contig of 16093 bp in length
            *    42617    42716: gap of unknown length
            *    42717    59134: contig of 16418 bp in length
            *    59135    59234: gap of unknown length
            *    59235    77937: contig of 18703 bp in length
            *    77938    78037: gap of unknown length
            *    78038    98933: contig of 20896 bp in length
            *    98934    99033: gap of unknown length
            *    99034   148068: contig of 49035 bp in length
            *   148069   148168: gap of unknown length
            *   148169   201416: contig of 53248 bp in length.
FEATURES             Location/Qualifiers
     source          1. .201416
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-740L19"
BASE COUNT    63751 a  38726 c  37375 g  60653 t    911 others
ORIGIN

  Query Match             88.2%;  Score 19.4;  DB 2;  Length 201416;
  Best Local Similarity   95.2%;  Pred. No. 1.6e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 AAGAAAAAATCTAGACAAGCA 21
          ||||||||||||||||| |||
Db 174985 AAGAAAAAATCTAGACAGGCA 174965


RESULT 9
AC022912
LOCUS       AC022912              218859 bp    DNA     linear   HTG 26-MAY-2000
DEFINITION  Homo sapiens clone RP11-740L19, WORKING DRAFT SEQUENCE, 34
            unordered pieces.
ACCESSION   AC022912
VERSION     AC022912.3  GI:7596818
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 218859)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-740L19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 218859)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 19, 2000 this sequence version replaced gi:7158108.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6286
                Center clone name: 740_L_19
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 186472 bases at least Q40
                Consensus quality: 198582 bases at least Q30
                Consensus quality: 206779 bases at least Q20
                Insert size: 185000; agarose-fp
                Insert size: 215559; sum-of-contigs
                Quality coverage: 4.5 in Q20 bases; agarose-fp
                Quality coverage: 3.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 34 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1210: contig of 1210 bp in length
            *     1211 1310: gap of      100 bp
            *     1311     2361: contig of 1051 bp in length
            *     2362 2461: gap of      100 bp
            *     2462     3783: contig of 1322 bp in length
            *     3784 3883: gap of      100 bp
            *     3884     5069: contig of 1186 bp in length
            *     5070 5169: gap of      100 bp
            *     5170     6265: contig of 1096 bp in length
            *     6266 6365: gap of      100 bp
            *     6366     7846: contig of 1481 bp in length
            *     7847 7946: gap of      100 bp
            *     7947     8992: contig of 1046 bp in length
            *     8993 9092: gap of      100 bp
            *     9093    10112: contig of 1020 bp in length
            *    10113 10212: gap of      100 bp
            *    10213    11548: contig of 1336 bp in length
            *    11549 11648: gap of      100 bp
            *    11649    12946: contig of 1298 bp in length
            *    12947 13046: gap of      100 bp
            *    13047    14927: contig of 1881 bp in length
            *    14928 15027: gap of      100 bp
            *    15028    16366: contig of 1339 bp in length
            *    16367 16466: gap of      100 bp
            *    16467    17996: contig of 1530 bp in length
            *    17997 18096: gap of      100 bp
            *    18097    20861: contig of 2765 bp in length
            *    20862 20961: gap of      100 bp
            *    20962    22546: contig of 1585 bp in length
            *    22547 22646: gap of      100 bp
            *    22647    25390: contig of 2744 bp in length
            *    25391 25490: gap of      100 bp
            *    25491    28906: contig of 3416 bp in length
            *    28907 29006: gap of      100 bp
            *    29007    33315: contig of 4309 bp in length
            *    33316 33415: gap of      100 bp
            *    33416    35211: contig of 1796 bp in length
            *    35212 35311: gap of      100 bp
            *    35312    40323: contig of 5012 bp in length
            *    40324 40423: gap of      100 bp
            *    40424    44661: contig of 4238 bp in length
            *    44662 44761: gap of      100 bp
            *    44762    50183: contig of 5422 bp in length
            *    50184 50283: gap of      100 bp
            *    50284    56549: contig of 6266 bp in length
            *    56550 56649: gap of      100 bp
            *    56650    63560: contig of 6911 bp in length
            *    63561 63660: gap of      100 bp
            *    63661    73767: contig of 10107 bp in length
            *    73768 73867: gap of      100 bp
            *    73868    84412: contig of 10545 bp in length
            *    84413 84512: gap of      100 bp
            *    84513    95985: contig of 11473 bp in length
            *    95986 96085: gap of      100 bp
            *    96086   109032: contig of 12947 bp in length
            *   109033 109132: gap of      100 bp
            *   109133   120785: contig of 11653 bp in length
            *   120786 120885: gap of      100 bp
            *   120886   132423: contig of 11538 bp in length
            *   132424 132523: gap of      100 bp
            *   132524   151790: contig of 19267 bp in length
            *   151791 151890: gap of      100 bp
            *   151891   170984: contig of 19094 bp in length
            *   170985 171084: gap of      100 bp
            *   171085   193704: contig of 22620 bp in length
            *   193705 193804: gap of      100 bp
            *   193805   218859: contig of 25055 bp in length.
FEATURES             Location/Qualifiers
     source          1. .218859
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-740L19"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1210
                     /note="assembly_fragment"
     misc_feature    1311. .2361
                     /note="assembly_fragment"
     misc_feature    2462. .3783
                     /note="assembly_fragment"
     misc_feature    3884. .5069
                     /note="assembly_fragment"
     misc_feature    5170. .6265
                     /note="assembly_fragment"
     misc_feature    6366. .7846
                     /note="assembly_fragment"
     misc_feature    7947. .8992
                     /note="assembly_fragment"
     misc_feature    9093. .10112
```

```
                     /note="assembly_fragment"
     misc_feature    10213. .11548
                     /note="assembly_fragment"
     misc_feature    11649. .12946
                     /note="assembly_fragment"
     misc_feature    13047. .14927
                     /note="assembly_fragment"
     misc_feature    15028. .16366
                     /note="assembly_fragment"
     misc_feature    16467. .17996
                     /note="assembly_fragment"
     misc_feature    18097. .20861
                     /note="assembly_fragment"
     misc_feature    20962. .22546
                     /note="assembly_fragment"
     misc_feature    22647. .25390
                     /note="assembly_fragment"
     misc_feature    25491. .28906
                     /note="assembly_fragment"
     misc_feature    29007. .33315
                     /note="assembly_fragment"
     misc_feature    33416. .35211
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    35312. .40323
                     /note="assembly_fragment"
     misc_feature    40424. .44661
                     /note="assembly_fragment"
     misc_feature    44762. .50183
                     /note="assembly_fragment"
     misc_feature    50284. .56549
                     /note="assembly_fragment"
     misc_feature    56650. .63560
                     /note="assembly_fragment"
     misc_feature    63661. .73767
                     /note="assembly_fragment"
     misc_feature    73868. .84412
                     /note="assembly_fragment"
     misc_feature    84513. .95985
                     /note="assembly_fragment"
     misc_feature    96086. .109032
                     /note="assembly_fragment"
     misc_feature    109133. .120785
                     /note="assembly_fragment"

  Query Match             88.2%;  Score 19.4;  DB 2;  Length 218859;
  Best Local Similarity   95.2%;  Pred. No. 1.6e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCA 21
           ||||||||||||||||| |||
Db  146022 AAGAAAAAATCTAGACAGGCA 146042


RESULT 10
AC093623/c
LOCUS       AC093623              274349 bp    DNA     linear   HTG 07-SEP-2001
DEFINITION  Homo sapiens chromosome UNK clone CTD-2335D13, *** SEQUENCING IN
            PROGRESS ***, 49 unordered pieces.
ACCESSION   AC093623
VERSION     AC093623.1  GI:15487445
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 274349)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 274349)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT
            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_MS2335D13
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 1%
            Sequencing vector: plasmid; 99%
            Chemistry: Dye-primer ET; 1% of reads
            Chemistry: Dye-terminator Big Dye; 99% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 248392 bases at least Q40
            Consensus quality: 257485 bases at least Q30
            Consensus quality: 262193 bases at least Q20
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 49 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1389: contig of 1389 bp in length
            *     1390     1489: gap of unknown length
            *     1490     2556: contig of 1067 bp in length
            *     2557     2656: gap of unknown length
            *     2657     3865: contig of 1209 bp in length
            *     3866     3965: gap of unknown length
            *     3966     5038: contig of 1073 bp in length
            *     5039     5138: gap of unknown length
            *     5139     6762: contig of 1624 bp in length
            *     6763     6862: gap of unknown length
            *     6863     8029: contig of 1167 bp in length
            *     8030     8129: gap of unknown length
            *     8130     9245: contig of 1116 bp in length
            *     9246     9345: gap of unknown length
            *     9346    10441: contig of 1096 bp in length
            *    10442    10541: gap of unknown length
            *    10542    11788: contig of 1247 bp in length
            *    11789    11888: gap of unknown length
            *    11889    13219: contig of 1331 bp in length
            *    13220    13319: gap of unknown length
            *    13320    14328: contig of 1009 bp in length
            *    14329    14428: gap of unknown length
            *    14429    15578: contig of 1150 bp in length
            *    15579    15678: gap of unknown length
            *    15679    17242: contig of 1564 bp in length
            *    17243    17342: gap of unknown length
            *    17343    19056: contig of 1714 bp in length
            *    19057    19156: gap of unknown length
            *    19157    20355: contig of 1199 bp in length
            *    20356    20455: gap of unknown length
            *    20456    21888: contig of 1433 bp in length
            *    21889    21988: gap of unknown length
            *    21989    23748: contig of 1760 bp in length
            *    23749    23848: gap of unknown length
            *    23849    25791: contig of 1943 bp in length
            *    25792    25891: gap of unknown length
            *    25892    29014: contig of 3123 bp in length
            *    29015    29114: gap of unknown length
            *    29115    30653: contig of 1539 bp in length
            *    30654    30753: gap of unknown length
            *    30754    32579: contig of 1826 bp in length
            *    32580    32679: gap of unknown length
            *    32680    34693: contig of 2014 bp in length
            *    34694    34793: gap of unknown length
```

```
          *    34794    36794: contig of 2001 bp in length
          *    36795    36894: gap of unknown length
          *    36895    39286: contig of 2392 bp in length
          *    39287    39386: gap of unknown length
          *    39387    42070: contig of 2684 bp in length
          *    42071    42170: gap of unknown length
          *    42171    45523: contig of 3353 bp in length
          *    45524    45623: gap of unknown length
          *    45624    47928: contig of 2305 bp in length
          *    47929    48028: gap of unknown length
          *    48029    50378: contig of 2350 bp in length
          *    50379    50478: gap of unknown length
          *    50479    52145: contig of 1667 bp in length
          *    52146    52245: gap of unknown length
          *    52246    54934: contig of 2689 bp in length
          *    54935    55034: gap of unknown length
          *    55035    57539: contig of 2505 bp in length
          *    57540    57639: gap of unknown length
          *    57640    60533: contig of 2894 bp in length
          *    60534    60633: gap of unknown length
          *    60634    63866: contig of 3233 bp in length
          *    63867    63966: gap of unknown length
          *    63967    67394: contig of 3428 bp in length
          *    67395    67494: gap of unknown length
          *    67495    71251: contig of 3757 bp in length
          *    71252    71351: gap of unknown length
          *    71352    75513: contig of 4162 bp in length
          *    75514    75613: gap of unknown length
          *    75614    79500: contig of 3887 bp in length
          *    79501    79600: gap of unknown length
          *    79601    85955: contig of 6355 bp in length
          *    85956    86055: gap of unknown length
          *    86056    90896: contig of 4841 bp in length
          *    90897    90996: gap of unknown length
          *    90997    95402: contig of 4406 bp in length
          *    95403    95502: gap of unknown length
          *    95503   100307: contig of 4805 bp in length
          *   100308   100407: gap of unknown length
          *   100408   105113: contig of 4706 bp in length
          *   105114   105213: gap of unknown length
          *   105214   109576: contig of 4363 bp in length
          *   109577   109676: gap of unknown length
          *   109677   114722: contig of 5046 bp in length
          *   114723   114822: gap of unknown length
          *   114823   119838: contig of 5016 bp in length
          *   119839   119938: gap of unknown length
          *   119939   128278: contig of 8340 bp in length
          *   128279   128378: gap of unknown length
          *   128379   137203: contig of 8825 bp in length
          *   137204   137303: gap of unknown length
          *   137304   170951: contig of 33648 bp in length
          *   170952   171051: gap of unknown length
          *   171052   274349: contig of 103298 bp in length.
FEATURES             Location/Qualifiers
     source          1. .274349
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="UNK"
                     /clone="CTD-2335D13"
     misc_feature    1. .1389
                     /note="assembly_name:Contig18"
     misc_feature    1490. .2556
                     /note="assembly_name:Contig22"
     misc_feature    2657. .3865
                     /note="assembly_name:Contig24"
     misc_feature    3966. .5038
                     /note="assembly_name:Contig27"
     misc_feature    5139. .6762
                     /note="assembly_name:Contig28"
     misc_feature    6863. .8029
                     /note="assembly_name:Contig30"
     misc_feature    8130. .9245
                     /note="assembly_name:Contig31"
     misc_feature    9346. .10441
                     /note="assembly_name:Contig32"
     misc_feature    10542. .11788
                     /note="assembly_name:Contig34"
     misc_feature    11889. .13219
                     /note="assembly_name:Contig35"
     misc_feature    13320. .14328
                     /note="assembly_name:Contig36"
     misc_feature    14429. .15578
                     /note="assembly_name:Contig37"
     misc_feature    15679. .17242
                     /note="assembly_name:Contig39"
     misc_feature    17343. .19056
                     /note="assembly_name:Contig41"
     misc_feature    19157. .20355
                     /note="assembly_name:Contig42"
     misc_feature    20456. .21888
                     /note="assembly_name:Contig43"
     misc_feature    21989. .23748
                     /note="assembly_name:Contig44"
     misc_feature    23849. .25791
                     /note="assembly_name:Contig45"
     misc_feature    25892. .29014
                     /note="assembly_name:Contig46"
     misc_feature    29115. .30653
                     /note="assembly_name:Contig47"
     misc_feature    30754. .32579
                     /note="assembly_name:Contig49"
     misc_feature    32680. .34693
                     /note="assembly_name:Contig50"
     misc_feature    34794. .36794
                     /note="assembly_name:Contig51"
     misc_feature    36895. .39286
                     /note="assembly_name:Contig52"
     misc_feature    39387. .42070
                     /note="assembly_name:Contig53"
     misc_feature    42171. .45523
                     /note="assembly_name:Contig54"

  Query Match             88.2%;  Score 19.4;  DB 2;  Length 274349;
  Best Local Similarity   95.2%;  Pred. No. 1.5e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCA 21
           ||||||||||||||||| |||
Db 220758 AAGAAAAAATCTAGACAGGCA 220738


RESULT 11
AP005572/c
LOCUS       AP005572              113940 bp    DNA     linear   HTG 24-JUL-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 9 clone
            OJ1506_A04, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION   AP005572
VERSION     AP005572.1  GI:21952940
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
            clone:OJ1506_A04.
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T., Hattori,M., Sakaki,Y. and Katayose,Y.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 9, BAC
            clone:OJ1506_A04
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 113940)
  AUTHORS   Sasaki,T., Matsumoto,T., Hattori,M., Sakaki,Y. and Katayose,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
```

```
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .113940
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="9"
                     /clone="OJ1506_A04"
BASE COUNT    33810 a  24133 c  23762 g  32028 t    207 others
ORIGIN

  Query Match             86.4%;  Score 19;  DB 2;  Length 113940;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AAGAAAAAATCTAGACAAG 19
          ||||||||||||||||||||
Db 112166 AAGAAAAAATCTAGACAAG 112148


RESULT 12
AF012746/c
LOCUS       AF012746               2052 bp    mRNA    linear   VRT 24-OCT-2000
DEFINITION  Danio rerio olfactory receptor protein 13.1 mRNA, complete cds.
ACCESSION   AF012746 U42397
VERSION     AF012746.1  GI:2331258
KEYWORDS    .
SOURCE      Danio rerio.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 2052)
  AUTHORS   Barth,A.L., Justice,N.J. and Ngai,J.
  TITLE     Asynchronous onset of odorant receptor expression in the developing
            zebrafish olfactory system
  JOURNAL   Neuron 16 (1), 23-34 (1996)
  MEDLINE   96158919
   PUBMED   8562087
REFERENCE   2  (bases 1 to 2052)
  AUTHORS   Barth,A.L., Dugas,J.C. and Ngai,J.
  TITLE     Noncoordinate expression of odorant receptor genes tightly linked
            in the zebrafish genome
  JOURNAL   Neuron 19 (2), 359-369 (1997)
  MEDLINE   97436752
   PUBMED   9292725
REFERENCE   3  (bases 1 to 2052)
  AUTHORS   Barth,A.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1995) Molecular and Cellular Biology, University
            of California, Berkeley, Rm. 265 LSA, Berkeley, CA 94720, USA
REFERENCE   4  (bases 1 to 2052)
  AUTHORS   Barth,A.L., Dugas,J.C. and Ngai,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-1997) Molecular and Cellular Biology, University
            of California, Berkeley, Rm. 265 LSA, Berkeley, CA 94720, USA
  REMARK    Sequence updated by submitter
COMMENT     On Oct 24, 2000 this sequence version replaced gi:1151128.
FEATURES             Location/Qualifiers
     source          1. .2052
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /tissue_type="olfactory"
     CDS             17. .1003
                     /note="previously named olfactory receptor protein 13"
                     /codon_start=1
                     /product="olfactory receptor protein 13.1"
                     /protein_id="AAC60253.1"
                     /db_xref="GI:2331259"
                     /translation="MSAGNISFVKEFFIVGFPGLQKGYFDSVAALMLFVFVCILAGNG
                     TFLILFSLEKSLHKPMYYVILNLVASDILFSTTTLPKIIARYLFQDGGISFMGCFIQM
                     YFVHYFGTVNAFILAIMAFDRYVAICNPLRYPNIVTKWNIFYLCLISWVIANVTPLMM
                     VIRAYPLPYCAENTIIQCYCDHISITSLACTNRAPYSIPAFVLAMVALLTPLAFIVFS
                     YCAIILAVLRISSTQARLKTFSTCSPQLIIIALYFLPRCFIYLSSNIGIYFSTDLRLA
                     IIMMYSLFPPMFNPLIYCLRTKDVKECLVKRLKISTNVKFNLTKVDISSVSS"
BASE COUNT      638 a    319 c    293 g    802 t
ORIGIN

  Query Match             85.5%;  Score 18.8;  DB 5;  Length 2052;
  Best Local Similarity   90.9%;  Pred. No. 6e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 AAGAAAAAATCTAGACAAGCAA 22
          |||||||||| ||| ||||||
Db   1472 AAGAAAAAATCAAGAGAAGCAA 1451


RESULT 13
SCYPL252C/c
LOCUS       SCYPL252C              2209 bp    DNA     linear   PLN 07-AUG-1997
DEFINITION  S.cerevisiae chromosome XVI reading frame ORF YPL252c.
ACCESSION   Z73608 U00094
VERSION     Z73608.1  GI:1370516
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2209)
  AUTHORS   Pohl,T.M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2209)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1996) Data collected by MIPS on behalf of the
            European yeast chromosome XVI sequencing project. MIPS at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .2209
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /chromosome="XVI"
     CDS             complement(<1. .907)
                     /note="ORF YPL253c"
                     /codon_start=1
                     /protein_id="CAA97977.1"
                     /db_xref="GI:2326844"
                     /db_xref="SPTREMBL:Q12045"
                     /translation="MASQQNKHAFLSKNRIFHNPDNVSSSKSRNLMDITNTTNTMNGS
                     RPSSMKSSLALPPVKDSFPSVSRSASLNINMSKIKDLKDRQDKIRFQRHTLRTQLIEC
                     EREIKTIKFRDLNKSRFELYKKKSKQAKYLKQVRDLTQNLNSKDGERADLIKKNKSAL
                     ATLQAELDQNLILKRQESQELYNNKLIFWENELQIMENVEPDHEITEEISQLKKTLQE
                     LNINWANLQKQNLERQVNHESQLRKDFIAFKEAKLKSMENLTNKHRELLDQIATLQSE
                     SEKLHKEIMDIDRQAEYSEQNISEIN"
     CDS             complement(1264. .1782)
                     /note="ORF YPL252c"
                     /codon_start=1
                     /protein_id="CAA97975.1"
                     /db_xref="GI:1370517"
                     /db_xref="SPTREMBL:Q12184"
                     /translation="MLKIVTRAGHTARISNIAAHLLRTSPSLLTRTTTTTRFLPFSTS
                     SFLNHGHLKKPKPGEELKITFILKDGSQKTYEVCEGETILDIAQGHNLDMEGACGGSC
```

```
                 ACSTCHVIVDPDYYDALPEPEDDENDMLDLAYGLTETSRLGCQIKMSKDIDGIRVALP
                 QMTRNVNNNDFS"
     CDS         1526. .1828
                 /note="ORF YPL251w"
                 /codon_start=1
                 /protein_id="CAA97976.1"
                 /db_xref="GI:1370518"
                 /db_xref="SPTREMBL:Q12058"
                 /translation="MSRMVSPSQTSYVFWEPSFRIKVIFSSSPGFGFFKWPWFKNEDV
                 EKGRNLVVVVVVRVSRDGEVRKRCAAMFDILAVCPARVTIFSIGGGLSIDTRSAYFV"
BASE COUNT      600 a    362 c    474 g    773 t
ORIGIN

  Query Match           85.5%;  Score 18.8;  DB 8;  Length 2209;
  Best Local Similarity  90.9%;  Pred. No. 6e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           |||||||||||  | ||||||||
Db     541 AAGAAAAAATCCAAACAAGCAA 520


RESULT 14
SCYPL253C/c
LOCUS       SCYPL253C               2411 bp    DNA     linear   PLN 11-AUG-1997
DEFINITION  S.cerevisiae chromosome XVI reading frame ORF YPL253c.
ACCESSION   Z73609 U00094
VERSION     Z73609.1  GI:1370519
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2411)
  AUTHORS   Pohl,T.M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2411)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1996) Data collected by MIPS on behalf of the
            European yeast chromosome XVI sequencing project. MIPS at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .2411
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /chromosome="XVI"
     CDS             complement(112. .2055)
                     /note="ORF YPL253c"
                     /codon_start=1
                     /protein_id="CAA97978.1"
                     /db_xref="GI:1370520"
                     /db_xref="SPTREMBL:Q12045"
                     /translation="MASQQNKHAFLSKNRIFHNPDNVSSSKSRNLMDITNTTNTMNGS
                     RPSSMKSSLALPPVKDSFPSVSRSASLNINMSKIKDLKDRQDKIRFQRHTLRTQLIEC
                     EREIKTIKFRDLNKSRFELYKKKSKQAKYLKQVRDLTQNLNSKDGERADLIKKNKSAL
                     ATLQAELDQNLILKRQESQELYNNKLIFWENELQIMENVEPDHEITEEISQLKKTLQE
                     LNINWANLQKQNLERQVNHESQLRKDFIAFKEAKLKSMENLTNKHRELLDQIATLQSE
                     SEKLHKEIMDIDRQAEYSEQNISEINENIKQLELANNPLISKSLQNSQDLEHLQNQME
                     NLKEMASKQEKFYNDTYNTVEKELLRSRRLENSIIEQKGTMRCYAYVMEQNLPENLLF
                     DYENGVITQGLSEHVYKFNRVIPHLKVSEDKFFTQEYSVYHDMCLNQKKNFNLISLST
                     TPHGSLRESLIKFLAEKDTIYQKQYVITLQFVFLSDDEFSQDMLLDYSHNDKDSIKLK
                     FEKHSISLDSKLVIIENGLEDLPLNFSCDEHPNLPHSGMGIIKVQFFPRDSKSDGNND
                     PVPVDFYFIELNNLKSIEQFDKSIFKKESCETPIALVLKKLISDTKSFFLLNLNDSKN
                     VNKLLTISEEVQTQLCKRKKKLT"
     misc_feature    1138. .1148
                     /note="ARS-consensus"
BASE COUNT      654 a    380 c    456 g    921 t
ORIGIN

  Query Match           85.5%;  Score 18.8;  DB 8;  Length 2411;
  Best Local Similarity  90.9%;  Pred. No. 5.9e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           |||||||||||  | ||||||||
Db    1689 AAGAAAAAATCCAAACAAGCAA 1668


RESULT 15
SC38KCXVI/c
LOCUS       SC38KCXVI               37808 bp    DNA     linear   PLN 21-OCT-1996
DEFINITION  S.cerevisiae DNA (chromosome XVI; 38 kb).
ACCESSION   Z67751
VERSION     Z67751.1  GI:1061234
KEYWORDS    gal4 protein; H(+)-transporting ATPase; HSP82 protein; HSP90
            protein; SRP68 protein; SUI3 protein.
SOURCE      Saccharomyces cerevisiae.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 37808)
  AUTHORS   Pohl,T.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-1995) Thomas M. Pohl, GATC GmbH,
            Fritz-Arnold-Str. 23, Konstanz, 78467, Germany
REFERENCE   2  (bases 1 to 37808)
  AUTHORS   Pohl,T.M.
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .37808
                     /organism="Saccharomyces cerevisiae"
                     /strain="alphaS288C"
                     /db_xref="taxon:4932"
                     /chromosome="XVI"
                     /clone_lib="cosmid library (A. Goffeau, Belgium)"
     CDS             870. .2336
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAA91590.1"
                     /db_xref="GI:1061235"
                     /db_xref="SWISS-PROT:Q12060"
                     /translation="MSAIQSPAPKPLQPTYPAASPASTNAYMKPGLIGSPAVSNHTEP
                     NNGNNETAEPQGPNQRIDLGAMIEELTSLLGKESWTKYAQIISLFILGKLSRKELSNE
                     LELVFSPSAASLEKSNTNHHHSLVRLHNQLLLGIFANSLRENPLGRNGNESSWGFGNG
                     SNNPNNKLKRINKHNSQIEVYKKIVMSLPLNDRNRLKMITKEAGKRGFIFCSVFQARL
                     NNIPKIPIVTNPESLKRVKSNNLKTPLEWSQDIMNGFNVPLASESHSLPDTDSFYLRM
                     VGIAREHGLVGTVDARCVELISLALDQYLKNIIEFTIDTVRYRRKKYSDYYDLNESGL
                     YKSVSEMAADKRDAKIKQLDDDKNEDECADEAKSINNGNNSSKDDIGDISMSSITKAG
                     EAVNEELHENRTISLTNEDIYDSLSIFPNLVEPSGSYYALTNLGLVNDDELVDMKSNI
                     DDLPDFLNEKPTFTPLDERNVGTRHELNWLIKGILTED"
     CDS             complement(2448. .4391)
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAA91591.1"
                     /db_xref="GI:1061236"
                     /db_xref="SPTREMBL:Q12045"
                     /translation="MASQQNKHAFLSKNRIFHNPDNVSSSKSRNLMDITNTTNTMNGS
                     RPSSMKSSLALPPVKDSFPSVSRSASLNINMSKIKDLKDRQDKIRFQRHTLRTQLIEC
                     EREIKTIKFRDLNKSRFELYKKKSKQAKYLKQVRDLTQNLNSKDGERADLIKKNKSAL
                     ATLQAELDQNLILKRQESQELYNNKLIFWENELQIMENVEPDHEITEEISQLKKTLQE
                     LNINWANLQKQNLERQVNHESQLRKDFIAFKEAKLKSMENLTNKHRELLDQIATLQSE
                     SEKLHKEIMDIDRQAEYSEQNISEINENIKQLELANNPLISKSLQNSQDLEHLQNQME
                     NLKEMASKQEKFYNDTYNTVEKELLRSRRLENSIIEQKGTMRCYAYVMEQNLPENLLF
                     DYENGVITQGLSEHVYKFNRVIPHLKVSEDKFFTQEYSVYHDMCLNQKKNFNLISLST
                     TPHGSLRESLIKFLAEKDTIYQKQYVITLQFVFLSDDEFSQDMLLDYSHNDKDSIKLK
                     FEKHSISLDSKLVIIENGLEDLPLNFSCDEHPNLPHSGMGIIKVQFFPRDSKSDGNND
                     PVPVDFYFIELNNLKSIEQFDKSIFKKESCETPIALVLKKLISDTKSFFLLNLNDSKN
                     VNKLLTISEEVQTQLCKRKKKLT"
     CDS             complement(4748. .5266)
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAA91592.1"
                     /db_xref="GI:1061237"
                     /db_xref="SPTREMBL:Q12184"
```

/translation="MLKIVTRAGHTARISNIAAHLLRTSPSLLTRTTTTTRFLPFSTS
SFLNHGHLKKPKPGEELKITFILKDGSQKTYEVCEGETILDIAQGHNLDMEGACGGSC
ACSTCHVIVDPDYYDALPEPEDDENDMLDLAYGLTETSRLGCQIKMSKDIDGIRVALP
QMTRNVNNNDFS"

CDS 5010. .5312
/codon_start=1
/product="putative protein"
/protein_id="CAA91593.1"
/db_xref="GI:1061238"
/db_xref="SPTREMBL:Q12058"
/translation="MSRMVSPSQTSYVFWEPSFRIKVIFSSSPGFGFFKWPWFKNEDV
EKGRNLVVVVVVRVSRDGEVRKRCAAMFDILAVCPARVTIFSIGGGLSIDTRSAYFV"

CDS complement(5694. .6104)
/codon_start=1
/product="putative protein"
/protein_id="CAA91594.1"
/db_xref="GI:1061239"
/db_xref="SPTREMBL:Q12048"
/translation="MSSVESSPISRYEDEVFPLSFSNVAFEPPMLSHSPDRSTYADDF
SQSYQQELLTFPLSYPIVDESECTHTKDKTDSNIITSTEDDCMFDMEFNGNAASAVAA
ASKESNSASGFAFASNDAFANVAQQNYRLWLSSV"

CDS complement(8054. .10738)
/codon_start=1
/product="putative protein"
/protein_id="CAA91595.1"
/db_xref="GI:1061240"
/db_xref="SPTREMBL:Q12344"
/translation="MSSDKSIEKNTDTIASEVHEGDNHSNNLGSMEEEIKSTPSDQYE
EIAIIPTEPLHSDKELNDKQQSLGHEAPTNVSREEPIGISGDEDTQITEQNVNEQRQE
TREPSSEIDLNEPLDVEKDVTTDVQAPNGLNIEKEYDAVKENEKVYADTKEVVSSPEN
REVTGKNSGGEKSSSSKFLDDESGTTTAANANDISISSEVTPERSSENDNNQIHITNE
VAAGINLNENKEQKAAIEDGPVTAENLSSETARKVPPIPTQIINEKGDNSSENEVSAI
PTTSSPPLPPRQNVATSTSPKLPPRGKQREQPPKTKNAVPPPLEEEMKSEKFRKNFEE
TKRNSYHHVPLTGSKTAQLESTAEINLIASRYRKTSHHLNKEGEETRESLQEGQSFLK
STFTSFLENLSEYNEVENVNEEDREMFKIDWSFWTQVVNDYATVASNEPENLEAHVTN
GIPPQIRGIIWQLMANSKSREMEDIYETLLDTECLHEATIRRDLRRTKFVAEDKMESL
YKVIKVYSVYDPDVGYTQGMGFIAAPLLINCENEAESFGLLVGLMKNYGLRELFLPGM
PGLMLMLYQFDRLLEEHSPSLYNRLIREGISSTMYATQWFLTFFAYKFPLEFVLRIFD
IVFVEGIEVLLKFAVNLMLKNEETLVKLRFDELLDFLKDELFNYYLMGNQDDASVVQM
GLSNGNSFKGNDDGTFSYNVDLFVHDAMTGVYITPLTLRRYRGEYVEIHEKEQKKEDH
YESLRIQNHQLQREAQKLEHDYSILNKENISAANELIQNRLNMEMLLDEKNDLINTIT
DIKSQIEEEIRKQNLPNPDASLPKADLREDLERTISRNNEVMRENGQLEERITELQAE
IDELININKEQVSTASLLERDSKAKGRKGWTGFKKVFK"

CDS complement(11096. .13741)
/codon_start=1
/product="GAL4"
/protein_id="CAA91596.1"
/db_xref="GI:1061241"
/db_xref="SWISS-PROT:P04386"
/translation="MKLLSSIEQACDICRLKKLKCSKEKPKCAKCLKNNWECRYSPKT
KRSPLTRAHLTEVESRLERLEQLFLLIFPREDLDMILKMDSLQDIKALLTGLFVQDNV
NKDAVTDRLASVETDMPLTLRQHRISATSSSEESSNKGQRQLTVSIDSAAHHDNSTIP
LDFMPRDALHGFDWSEEDDMSDGLPFLKTDPNNNGFFGDGSLLCILRSIGFKPENYTN
SNVNRLPTMITDRYTLASRSTTSRLLQSYLNNFHPYCPIVHSPTLMMLYNNQIEIASK
DQWQILFNCILAIGAWCIEGESTDIDVFYYQNAKSHLTSKVFESGSIILVTALHLLSR
YTQWRQKTNTSYNFHSFSIRMAISLGLNRDLPSSFSDSSILEQRRRIWWSVYSWEIQL
SLLYGRSIQLSQNTISFPSSVDDVQRTTTGPTIYHGIIETARLLQVFTKIYELDKTVT
AEKSPICAKKCLMICNEIEEVSRQAPKFLQMDISTTALTNLLKEHPWLSFTRFELKWK
QLSLIIYVLRDFFTNFTQKKSQLEQDQNDHQSYEVKRCSIMLSDAAQRTVMSVSSYMD
NHNVTPYFAWNCSYYLFNAVLVPIKTLLSNSKSNAENNETAQLLQQINTVLMLLKKLA
TFKIQTCEKYIQVLEEVCAPFLLSQCAIPLPHISYNNSNGSAIKNIVGSATIAQYPTL
PEENVNNISVKYVSPGSVGPSPVPLKSGASFSDLVKLLSNRPPSRNSPVTIPRSTPSH
RSVTPFLGQQQQLQSLVPLTPSALFGGANFNQSGNIADSSLSFTFTNSSNGPNLITTQ
TNSQALSQPIASSNVHDNFMNNEITASKIDDGNNSKPLSPGWTDQTAYNAFGITTGMF
NTTTMDDVYNYLFDDEDTPPNPKKE"

CDS complement(14010. .15581)
/codon_start=1
/product="putative protein"
/protein_id="CAA91597.1"
/db_xref="GI:1061242"
/db_xref="SPTREMBL:Q12523"
/translation="MDPFHNGNKRSSISFGSSQRQPYNKNNYLSGTNGPSSAAQDQGR
GPSPFGMSGNTTNGGNSKRNSGCDLSATYYASRSPMYSPLDFSPPVFSPNHSQLQQAR
GYAANIPVVSNLMNPSMASVCEYQSHYPLFGLDWSADDYVCLGSYKEDSRNKLQVLHS
NDLLSWESVVDADVVYPVSKIQWVPSQLYPRKLATCSDSLRIWSVSPEERQFQEQINL
SLCKYNRQHPANPAAADDMKVIGTFPPITSFDWNTVDTNLIISSSIDTTCIVWDLQSS
HYVKTQLIAHDSEVFDVRFLTKSTQLFASCGGDGSVRVFDLRSLAHSTIIYEPPSSSV
SGATAGTITPSSKGSDALLRLEPSPYDPNVLATFAADSNKIIILDMRNPESPILNLQG
HGSSVNGIKWHPTKRNVLLSCGDDCQVLYWDLNSSFMEINAAGSKSPSIHGTSLEDPD
GDTEMTDGGAGSGLNEDPLSLNNNSKQVCKTLETPNMMYANKTQEINNIAWRPQRGDW
FGCVSGKKFQNVRVL"

CDS complement(15894. .16682)
/codon_start=1
/product="putative protein"
/protein_id="CAA91598.1"
/db_xref="GI:1061243"
/db_xref="SPTREMBL:Q12270"
/translation="MNWKSYVFPGGHPPAALTTGLVVFLTAIYLLSFIFALREDLSLA
PESLFKLQMSRLSLYPLIHLSLPHLLFNVLAIWAPLNLFEETHGTVYTGVFLNLSALF
AGILYCLLGKLLYPEALVAGASGWCFTLFAYYSFKESQIRPRTRIFRTDYSIPTLYTP
LVLLVAIAVVIPGSSFWGHFFGLCVGYAIGYKESWFNKITPPGWIITKIEKSLDGLIR
LIPWGIKYYRDEDIDRTKDYEPLMSTETPLPLHNDNSGTVLGTA"

CDS 16971. .18335
/codon_start=1
/product="putative protein"
/protein_id="CAA91599.1"
/db_xref="GI:1061244"
/db_xref="SPTREMBL:Q12179"
/translation="MHGPTSKAISRNVRSVKRPRRAPRPVVSTQAMNKLSNVTLSAEQ
EKLRERVLSFMRSNLSQYKSDWKHPAMFVIQGDAGTGKSVILNSLFNEIQKLSQFSPS
SEDILHGTHNYLVVNHPEMLKLYIRISDSFKYISKSSLERPTSLINNLQKRKVMADVV
IVDEAHLLATSKDAFKRFYGENHLKDLMSLCKVLVLVYDDKQALRMGSYWDEGSNNGA
TLKDFYNEIPPKSRDWYTLKQQFRVAAPQNVLNWIDQISVAGKIPPIESVLSKGNADC
ADDKIKNFDFKIWDDCGAMYEAIKEKDRQYGQCRMLSTYDFPYRLDGKDYYVECGDNF
KVRWDRYTPREVTPWSERCDTIDEVGSVYTIQGFDLNYAGVILGRSIGYDAANDCIKL
RPELYDDRAGFTKKKNIHNAEDVKQKIIMNSINVLLTRGVRGLYVYAYDPELRERLLR
PSKK"

CDS complement(18399. .19418)
/codon_start=1
/product="putative protein"
/protein_id="CAA91600.1"
/db_xref="GI:1061245"
/db_xref="SPTREMBL:Q12520"
/translation="MAGSTSSLVICAIGIYATFLTWALVQEPLATRTWPNSMGKFQFP
NVISLIQASVAMMMGYLYLNWKKVEYPPRKMIKDHWKQLMLISFTQSSSGPLATTSLK
HVDYLTYMLAKSCKMIPVLLVHLLLYRTPIASQKKVVALLVSLGVTIFTIGGNDGKKL
KRSFNESGNDNKLQGFGLLFSSLFLDGLTNATQDKLLKANKAKEKGKQTLITGAHLMF
TLNLFVILWNILYFIVIDCKQWDNAVSVLTMDPQVWGYLMLYSFCGAMGQCFIFYTLE
QFGSLVLIMITVTRKMVSMILSIIVFGKSVRFQQWVGMFIVFGGITWEALNKKKANIP
KAKSA"

CDS 19902. .21701
/codon_start=1
/product="SRP68"
/protein_id="CAA91601.1"
/db_xref="GI:1061246"
/db_xref="SWISS-PROT:P38687"
/translation="MVAYSPIIATYGNRAEQFLETDSDFAKYHAKLNKKLQHLRSRCH
LVTKDTKKYSSKNKYGEINSEDYDNKTKLIGVLILLHAERDLALAETLKLRARQRGKL
KKSEEKVLSTRLKKACKTADKLVNVTQNEQQWITRAQYLAFAKLVHSEYLINGKRFKR
KDNAKISNNLALVFAALEHLKNLSLLAEEVVDNIVNKYQYSLKQYAGNLITTPEINNF
IVERVQSDENKDDELVKLLLDNGFNMKKITTSTEDQKVTTNINWRSFNAKIIDAEVAQ
FLEQGLSIHPTQITQYTQRLSKLEKALDRHEFFIANHDDQDDIDEMVENSSENNQIIL
AYIKYNILLTSISRERDLFTHLWNQWLKLNTSLPSKLTKYKEMERIVKNLTKYLSDIM
ELPGVYSDDELLSQLDLCKLYFQLFLNTGCLSVLYQSKGRYMEALALYVDAYRRLENK
LSEIESLDEILLPANLLSLNSVRSLQKRIENGGNSVITLAEYEKRNHGGSLGKYDLTV
IEKLDSKKILPTDIQLKNLFPLKPKMLPIPSKPTLFDLAFNYITYDKQEPSASQVKDS

Query Match 85.5%; Score 18.8; DB 8; Length 37808;
Best Local Similarity 90.9%; Pred. No. 3.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAGCAA 22
||||||||||| | ||||||||
Db 4025 AAGAAAAAATCCAAACAAGCAA 4004

Search completed: March 17, 2003, 11:31:31
Job time : 548.495 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:42:12 ; Search time 126.086 Seconds
(without alignments)
392.938 Million cell updates/sec

Title: US-09-836-439-4
Perfect score: 22
Sequence: 1 aagaaaaaatctagacaagcaa 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	17.8	80.9	2483	24	ABQ69157	Listeria monocytog
c	2	17.4	79.1	370	22	ABA12703	Human nervous syst
c	3	17.4	79.1	412	22	ABA18135	Human nervous syst
c	4	17.4	79.1	412	22	ABA18136	Human nervous syst
c	5	17.2	78.2	434	23	ABV37860	Human prostate exp
c	6	17.2	78.2	507	23	ABV58334	Human prostate exp
	7	17.2	78.2	674	22	AAS30250	DNA encoding rena
	8	17.2	78.2	674	22	AAS30251	DNA encoding rena
	9	17.2	78.2	674	22	AAS26974	Human genomic DNA
	10	17.2	78.2	674	22	AAS26975	Human genomic DNA
	11	17.2	78.2	674	22	AAS33485	DNA encoding human
	12	17.2	78.2	674	22	AAS33486	DNA encoding human
c	13	17.2	78.2	782	24	AAD37775	Extented sequence
c	14	17.2	78.2	786	24	ABQ73702	Human colon specif
c	15	17.2	78.2	831	24	ABQ43712	Oligonucleotide fo
	16	17.2	78.2	831	24	ABQ43713	Oligonucleotide fo
c	17	17.2	78.2	1372	23	ABV25096	Human prostate exp
c	18	17.2	78.2	1509	24	ABQ73703	Human colon specif
	19	17.2	78.2	1588	21	AAC35087	Arabidopsis thalia
c	20	17.2	78.2	1622	22	AAK83087	Human immune/haema
	21	17.2	78.2	2713	23	ABL24452	Drosophila melanog
c	22	17.2	78.2	2961	24	ABQ77508	Human cytokine rec
	23	17.2	78.2	2963	23	ABL22334	Drosophila melanog
c	24	17.2	78.2	3138	23	ABL24782	Drosophila melanog
	25	17.2	78.2	3729	22	AAF77688	Human wild-type Fc
	26	17.2	78.2	4081	23	ABL16178	Drosophila melanog
	27	17.2	78.2	4131	22	AAF77689	Human variant Fcep
	28	17.2	78.2	4258	22	AAK67062	Human immune/haema
	29	17.2	78.2	5552	21	AAA58309	pDIP/PyCSP.1 plasm
c	30	17.2	78.2	6716	24	ABL33783	Human immune syste
c	31	17.2	78.2	7202	22	ABA19579	Human nervous syst
c	32	17.2	78.2	7202	22	AAK71750	Human immune/haema
	33	17.2	78.2	9615	22	AAL36590	Human musculoskele
	34	17.2	78.2	11298	18	AAT86756	Human high affinit
	35	17.2	78.2	11298	19	AAV54661	Human beta subunit
	36	17.2	78.2	11298	21	AAF20937	Human high affinit
	37	17.2	78.2	11298	21	AAA34815	Human adenosine re
	38	17.2	78.2	11298	22	AAF92144	Human IGERB gene S
	39	17.2	78.2	11357	14	AAQ51024	Human FcERI beta c
	40	17.2	78.2	21742	21	AAF20938	Human high affinit
	41	17.2	78.2	21742	21	AAA34816	Human adenosine re
c	42	17.2	78.2	24512	22	ABA19578	Human nervous syst
c	43	17.2	78.2	24512	22	AAK67571	Human immune/haema
c	44	17.2	78.2	24512	22	AAK71749	Human immune/haema
	45	17.2	78.2	33030	22	AAF29337	Atopy related gene

ALIGNMENTS

RESULT 1
ABQ69157
ID ABQ69157 standard; DNA; 2483 BP.
XX
AC ABQ69157;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes 4b contig DNA sequence #1923.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
OS Listeria monocytogenes 4b.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,

PT antibodies and modulators -
XX
PS Claim 14; SEQ ID 1970; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2483 BP; 771 A; 651 C; 362 G; 699 T; 0 other;

Query Match 80.9%; Score 17.8; DB 24; Length 2483;
Best Local Similarity 90.5%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAGCA 21
||||||||||| | |||||||
Db 2200 AAGAAAAAATCCATACAAGCA 2220

RESULT 2
ABA12703/c
ID ABA12703 standard; cDNA; 370 BP.
XX
AC ABA12703;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 1710.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
DR P-PSDB; ABB16377.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Claim 1; SEQ ID NO 1710; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 370 BP; 130 A; 59 C; 60 G; 120 T; 1 other;

Query Match 79.1%; Score 17.4; DB 22; Length 370;
Best Local Similarity 94.7%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAG 19
||||||| ||||||||||
Db 223 AAGAAAATATCTAGACAAG 205

RESULT 3
ABA18135/c
ID ABA18135 standard; DNA; 412 BP.
XX
AC ABA18135;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 10466.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 10466; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 412 BP; 137 A; 69 C; 71 G; 135 T; 0 other;

Query Match 79.1%; Score 17.4; DB 22; Length 412;
Best Local Similarity 94.7%; Pred. No. 3.8e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAG 19
||||||| ||||||||||
Db 220 AAGAAAATATCTAGACAAG 202

RESULT 4
ABA18136/c
ID ABA18136 standard; DNA; 412 BP.
XX
AC ABA18136;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 10467.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.

PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT **Nucleic acids encoding 3224 human nervous system antigen polypeptides,**
PT **useful for preventing, diagnosing and/or treating nervous system**
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 10467; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 412 BP; 137 A; 69 C; 71 G; 135 T; 0 other;

Query Match 79.1%; Score 17.4; DB 22; Length 412;
Best Local Similarity 94.7%; Pred. No. 3.8e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 AAGAAAAAATCTAGACAAG 19
          ||||||| ||||||||||
Db    220 AAGAAAATATCTAGACAAG 202
```

RESULT 5
ABV37860/c
ID ABV37860 standard; cDNA; 434 BP.
XX
AC ABV37860;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 37851.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 7745-7746; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 434 BP; 130 A; 78 C; 76 G; 150 T; 0 other;

Query Match 78.2%; Score 17.2; DB 23; Length 434;
Best Local Similarity 86.4%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 AAGAAAAAATCTAGACAAGCAA 22
          |||||||||||| || ||| ||
Db    240 AAGAAAAAATCTTGAAAAGAAA 219
```

RESULT 6
ABV58334/c
ID ABV58334 standard; cDNA; 507 BP.
XX
AC ABV58334;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 58325.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX

PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 11199; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 507 BP; 179 A; 98 C; 93 G; 134 T; 3 other;

Query Match 78.2%; Score 17.2; DB 23; Length 507;
Best Local Similarity 86.4%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAGCAA 22
||||||| |||| |||||||
Db 296 AAGAAAAACACTAGTCAAGCAA 275

RESULT 7
AAS30250
ID AAS30250 standard; DNA; 674 BP.
XX
AC AAS30250;
XX
DT 21-NOV-2001 (first entry)
XX
DE DNA encoding renal and cardiovascular-associated protein, Seq ID 168.
XX
KW Human; antiinflammatory; neuroprotective; immunomodulator; vulnerary;
KW cardiovascular; cytostatic; nephrotropic; antianaemic; nephritis;
KW immunosuppressive; kidney disorder; renal failure; hypertension;
KW cardiovascular disorder; myocardial infarction; blood disorder; anaemia;
KW blood coagulation disorder; electrolyte imbalance disorder; cancer;
KW hyponatraemia; hyperkalaemia; neoplastic disorder; nephtoma;
KW autoimmune disease; inflammatory disease; reproductive system disorder;
KW endocrine disorder; neural activity; neurological disorder;
KW wound healing; respiratory disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200155328-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01359.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488787/53.
XX
PT New polynucleotides and polypeptides, useful for diagnosing, treating,
PT preventing or prognosing e.g. kidney, cardiovascular, blood,
PT electrolyte imbalance or neoplastic disorders, autoimmune diseases,
PT cancers -
XX
PS Claim 1; SEQ ID No 168; 506pp; English.
XX
CC The invention relates to novel nucleic acids and polypeptides useful for
CC diagnosing, treating, preventing and/or prognosing disorders related to
CC these polypeptides. The polynucleotides are especially useful in the
CC diagnosis, prognosis, prevention and/or treatment of diseases which
CC include kidney disorders (e.g. renal failure or nephritis),
CC cardiovascular disorders (e.g. hypertension or myocardial infarction),
CC blood disorders (e.g. anaemia or blood coagulation disorders),
CC electrolyte imbalance disorders (e.g. hyponatraemia or hyperkalaemia),
CC neoplastic disorders (e.g. nephtoma or renal cell cancer), autoimmune
CC diseases, cancers, inflammatory diseases, reproductive system
CC disorders, endocrine disorders, neural activity and neurological
CC disorders, wound healing and respiratory disorders. AAS30165-AAS30251
CC represent the novel human renal and cardiovascular-associated nucleic
CC acid sequences of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at:
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 674 BP; 144 A; 184 C; 145 G; 201 T; 0 other;

Query Match 78.2%; Score 17.2; DB 22; Length 674;
Best Local Similarity 86.4%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAGCAA 22
||||||||||| || |||| |||
Db 137 AAGAAAAAATTTAAACAAACAA 158

RESULT 8
AAS30251
ID AAS30251 standard; DNA; 674 BP.
XX
AC AAS30251;
XX
DT 21-NOV-2001 (first entry)
XX
DE DNA encoding renal and cardiovascular-associated protein, Seq ID 169.
XX
KW Human; antiinflammatory; neuroprotective; immunomodulator; vulnerary;
KW cardiovascular; cytostatic; nephrotropic; antianaemic; nephritis;
KW immunosuppressive; kidney disorder; renal failure; hypertension;
KW cardiovascular disorder; myocardial infarction; blood disorder; anaemia;
KW blood coagulation disorder; electrolyte imbalance disorder; cancer;
KW hyponatraemia; hyperkalaemia; neoplastic disorder; nephtoma;
KW autoimmune disease; inflammatory disease; reproductive system disorder;
KW endocrine disorder; neural activity; neurological disorder;
KW wound healing; respiratory disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200155328-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01359.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488787/53.
XX
PT New polynucleotides and polypeptides, useful for diagnosing, treating,
PT preventing or prognosing e.g. kidney, cardiovascular, blood,
PT electrolyte imbalance or neoplastic disorders, autoimmune diseases,
PT cancers -
XX
PS Claim 1; SEQ ID No 169; 506pp; English.
XX
CC The invention relates to novel nucleic acids and polypeptides useful for
CC diagnosing, treating, preventing and/or prognosing disorders related to
CC these polypeptides. The polynucleotides are especially useful in the
CC diagnosis, prognosis, prevention and/or treatment of diseases which
CC include kidney disorders (e.g. renal failure or nephritis),

CC cardiovascular disorders (e.g. hypertension or myocardial infarction),
CC blood disorders (e.g. anaemia or blood coagulation disorders),
CC electrolyte imbalance disorders (e.g. hyponatraemia or hyperkalaemia),
CC neoplastic disorders (e.g. nephtoma or renal cell cancer), autoimmune
CC diseases, cancers, inflammatory diseases, reproductive system
CC disorders, endocrine disorders, neural activity and neurological
CC disorders, wound healing and respiratory disorders. AAS30165-AAS30251
CC represent the novel human renal and cardiovascular-associated nucleic
CC acid sequences of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at:
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 674 BP; 144 A; 184 C; 145 G; 201 T; 0 other;

Query Match 78.2%; Score 17.2; DB 22; Length 674;
Best Local Similarity 86.4%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAGCAA 22
||||||||| || |||| |||
Db 137 AAGAAAAAATTTAAACAAACAA 158

RESULT 9
AAS26974
ID AAS26974 standard; DNA; 674 BP.
XX
AC AAS26974;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human genomic DNA encoding partial novel secreted protein, Seq ID 310.
XX
KW Human; immunosuppressive; antiarthritic; ds; antirheumatic;
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW vulnerary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155441-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01320.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-476222/51.
XX
PT Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, for treating blood clotting disorder,
PT haemophilia -
XX
PS Disclosure; SEQ ID No 310; 601pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence is a genomic DNA encoding a partial novel secreted protein of

Query Match 78.2%; Score 17.2; DB 22; Length 674;
Best Local Similarity 86.4%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAGCAA 22
|||||||||| || |||| |||
Db 137 AAGAAAAAATTTAAACAAACAA 158

RESULT 10
AAS26975
ID AAS26975 standard; DNA; 674 BP.
XX
AC AAS26975;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human genomic DNA encoding partial novel secreted protein, Seq ID 311.
XX
KW Human; immunosuppressive; antiarthritic; ds; antirheumatic;
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW vulnerary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155441-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01320.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-476222/51.
XX
PT Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, for treating blood clotting disorder,
PT haemophilia -
XX
PS Disclosure; SEQ ID No 311; 601pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence is a genomic DNA encoding a partial novel secreted protein of

```
  Query Match             78.2%;  Score 17.2;  DB 22;  Length 674;
  Best Local Similarity   86.4%;  Pred. No. 4.6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           |||||||||| || |||| |||
Db     137 AAGAAAAAATTTAAACAAACAA 158


RESULT 11
AAS33485
ID   AAS33485 standard; DNA; 674 BP.
XX
AC   AAS33485;
XX
DT   04-DEC-2001  (first entry)
XX
DE   DNA encoding human secreted protein, Seq ID No 768.
XX
KW   Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW   rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW   cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW   cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW   multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW   Gaucher's disease; neurological disease; cerebrovascular disorder;
KW   thrombosis; wound healing; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155326-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01347.
XX
PR   31-JAN-2000; 2000US-0179065.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-451931/48.
XX
PT   New nucleic acids and polypeptides, useful for diagnosing, preventing
PT   or treating medical conditions -
XX
PS   Disclosure; SEQ ID No 768; 753pp; English.
XX
CC   The invention relates to novel isolated nucleic acid molecules (I)
CC   encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC   treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC   goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC   the prevention, treatment and diagnosis of diseases associated with
CC   inappropriate expression of secreted proteins. (I) and complementary
CC   sequences may also be used as DNA probes in diagnostic assays (e.g.
CC   polymerase chain reactions (PCR)) to detect and quantitate the presence
CC   of similar nucleic acid sequences in samples, and so which patients may
CC   be in need of restorative therapy. (II) may also be used as antigens in
CC   the production of antibodies and in assays to identify modulators
CC   (agonists and antagonists) of the expression and activity of the secreted
CC   proteins. The anti-(II) antibodies and antagonists may also be used to
CC   down regulate expression and activity of (II). The anti-(II) antibodies
CC   may also be used as diagnostic agents for detecting the presence of (II)
CC   in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC   disorders include for example: immune/autoimmune diseases (e.g. HIV
CC   (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC   and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC   melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC   Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC   Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
CC   cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC   angina and thrombosis), infections caused by bacteria, viruses and
CC   fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC   agonists, antagonists and antibodies can also be used to promote wound
CC   healing, maintain organs before transplantation, and support cell culture
CC   of primary tissues. AAS33043-AAS33486 represent human secreted protein
CC   coding sequences, PCR primers, and related sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification but was obtained in electronic format directly from WIPO
CC   at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 674 BP; 144 A; 184 C; 145 G; 201 T; 0 other;

  Query Match             78.2%;  Score 17.2;  DB 22;  Length 674;
  Best Local Similarity   86.4%;  Pred. No. 4.6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           |||||||||| || |||| |||
Db     137 AAGAAAAAATTTAAACAAACAA 158


RESULT 12
AAS33486
ID   AAS33486 standard; DNA; 674 BP.
XX
AC   AAS33486;
XX
DT   04-DEC-2001  (first entry)
XX
DE   DNA encoding human secreted protein, Seq ID No 769.
XX
KW   Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW   rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW   cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW   cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW   multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW   Gaucher's disease; neurological disease; cerebrovascular disorder;
KW   thrombosis; wound healing; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155326-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01347.
XX
PR   31-JAN-2000; 2000US-0179065.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-451931/48.
XX
PT   New nucleic acids and polypeptides, useful for diagnosing, preventing
PT   or treating medical conditions -
XX
PS   Disclosure; SEQ ID No 769; 753pp; English.
XX
CC   The invention relates to novel isolated nucleic acid molecules (I)
CC   encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC   treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC   goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC   the prevention, treatment and diagnosis of diseases associated with
CC   inappropriate expression of secreted proteins. (I) and complementary
CC   sequences may also be used as DNA probes in diagnostic assays (e.g.
CC   polymerase chain reactions (PCR)) to detect and quantitate the presence
CC   of similar nucleic acid sequences in samples, and so which patients may
CC   be in need of restorative therapy. (II) may also be used as antigens in
CC   the production of antibodies and in assays to identify modulators
CC   (agonists and antagonists) of the expression and activity of the secreted
CC   proteins. The anti-(II) antibodies and antagonists may also be used to
CC   down regulate expression and activity of (II). The anti-(II) antibodies
```

CC may also be used as diagnostic agents for detecting the presence of (II)
CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC angina and thrombosis), infections caused by bacteria, viruses and
CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC agonists, antagonists and antibodies can also be used to promote wound
CC healing, maintain organs before transplantation, and support cell culture
CC of primary tissues. AAS33043-AAS33486 represent human secreted protein
CC coding sequences, PCR primers, and related sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification but was obtained in electronic format directly from WIPO
CC at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 674 BP; 144 A; 184 C; 145 G; 201 T; 0 other;

Query Match 78.2%; Score 17.2; DB 22; Length 674;
Best Local Similarity 86.4%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAGCAA 22
|||||||||| || |||| |||
Db 137 AAGAAAAAATTTAAACAAACAA 158

RESULT 13
AAD37775/c
ID AAD37775 standard; DNA; 782 BP.
XX
AC AAD37775;
XX
DT 27-AUG-2002 (first entry)
XX
DE Extented sequence for mouse IMX5_09.
XX
KW Inflammatory bowel disease; IBD; autoimmune disorder; arthritis; allergy;
KW haematopoietic cell; thrombolytic; blood coagulation disorder; nephritis;
KW asthma; organ rejection; graft-versus-host disease; inflammation; shock;
KW nerve disease; Alzheimer's disease; Parkinson's disease; antibacterial;
KW Huntington's disease; immunosuppressive; sepsis; nephrotropic; nootropic;
KW neuroprotective; anticonvulsant; gene therapy; mouse; ds.
XX
OS Mus musculus.
XX
PN WO200231116-A2.
XX
PD 18-APR-2002.
XX
PF 11-OCT-2001; 2001WO-US32176.
XX
PR 11-OCT-2000; 2000US-239712P.
XX
PA (DIGI-) DIGITAL GENE TECHNOLOGIES INC.
XX
PI Viney JL, Sims JE, Dubose RF, Baum PR, Hasel KW, Hilbush BS;
XX
DR WPI; 2002-426280/45.
XX
PT New polynucleotide associated with inflammatory bowel disease for
PT treating disorders of the immune system, nervous system, hematopoietic
PT cells and to modulate inflammation -
XX
PS Claim 1; Page 202-203; 214pp; English.
XX
CC The invention relates to an isolated polynucleotide associated with
CC inflammatory bowel disease (IBD). The invention is useful for
CC manufacturing a medicament for use in preventing, treating, modulating,
CC or ameliorating a medical condition which is IBD. The polypeptide and
CC polynucleotide are useful for treating disorders of the immune system
CC e.g. autoimmune disorders, deficiencies or disorders of haematopoietic
CC cells, to modulate haemostatic, or thrombolytic activity, treat blood
CC coagulation disorders, allergic reactions and conditions, such as asthma,
CC treat and/or prevent organ rejection or graft-versus-host disease and
CC modulate inflammation, including inflammation associated with infection,
CC shock, sepsis, arthritis and nephritis. The invention is useful to
CC differentiate, proliferate and attract cells, leading to the regeneration
CC of tissues and to treat central and peripheral nerve diseases e.g.
CC Alzheimer's disease, Parkinson's disease, and Huntington's disease. The
CC invention is useful in gene therapy. The present sequence is an extended
CC sequence for mouse IMX5_09 which is used in the exemplification of the
CC invention.
XX
SQ Sequence 782 BP; 212 A; 180 C; 138 G; 252 T; 0 other;

Query Match 78.2%; Score 17.2; DB 24; Length 782;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAGCAA 22
|| |||||||||| |||| |||
Db 393 AAAAAAAAATCTAAACAAACAA 372

RESULT 14
ABQ73702/c
ID ABQ73702 standard; cDNA; 786 BP.
XX
AC ABQ73702;
XX
DT 07-OCT-2002 (first entry)
XX
DE Human colon specific nucleic acid (CSNA) SEQ ID NO:8.
XX
KW Human; colon specific nucleic acid; colon specific polypeptide; CSP;
KW CSNA; colon specific gene; CSG; colon cancer; gene therapy; vaccine;
KW cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200248370-A2.
XX
PD 20-JUN-2002.
XX
PF 30-OCT-2001; 2001WO-US51341.
XX
PR 31-OCT-2000; 2000US-244717P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Sun Y, Recipon H, Ghosh MG, Liu C;
XX
DR WPI; 2002-583520/62.
XX
PT Colon specific polypeptides and polynucleotides useful for detecting,
PT diagnosing, monitoring, treating, staging and predicting cancers in
PT humans having cancer and non-cancerous colon disease -
XX
PS Claim 1; Page 153; 243pp; English.
XX
CC ABQ73695 to ABQ73841 represent human colon specific nucleic acid (CSNA)
CC sequences, and ABP51826 to ABP51928 represent human colon specific
CC polypeptide (CSP) sequences from the present invention. CSNA and CSP
CC sequences have cytostatic activity, and can be used in gene therapy,
CC antisense therapy and in vaccines. CSNA and CSP sequences can be used
CC for diagnosing and monitoring the presence and metastases of colon
CC cancer in a patient, by determining an amount of CSP or CSNA in a sample
CC of a patient, and comparing it to the amount of colon specific marker
CC in a normal control, where a difference in the amount of the nucleic
CC acid or the polypeptide in the sample compared to that of normal control
CC is associated with presence of colon cancer. CSP and CSNA sequences can
CC be used for producing engineered colon tissue for treatment and research.

CC CSNA sequences are useful for producing transgenic animals and cells
CC and also in gene therapy.
XX
SQ Sequence 786 BP; 256 A; 118 C; 140 G; 272 T; 0 other;

Query Match 78.2%; Score 17.2; DB 24; Length 786;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAGCAA 22
| || |||||||| ||||||||
Db 715 ATGAGAAAATCTAAACAAGCAA 694

RESULT 15
ABQ43712/c
ID ABQ43712 standard; DNA; 831 BP.
XX
AC ABQ43712;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 30303.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX

SQ Sequence 831 BP; 125 A; 99 C; 281 G; 326 T; 0 other;

Query Match 78.2%; Score 17.2; DB 24; Length 831;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAGCAA 22
|| |||||||||| |||| |||
Db 771 AAAAAAAAATCTAAACAACCAA 750

Search completed: March 17, 2003, 10:50:47
Job time : 128.253 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 10:23:12 ; Search time 777.688 Seconds
(without alignments)
458.154 Million cell updates/sec

Title: US-09-836-439-4
Perfect score: 22
Sequence: 1 aagaaaaaatctagacaagcaa 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	20.4	92.7	342	10	AW740409	AW740409 BRI10552
	2	20.4	92.7	342	12	BG360474	BG360474 BRI10646
	3	20.4	92.7	360	10	AW740410	AW740410 BRI10553
	4	18.8	85.5	224	10	BB016617	BB016617 BB016617
	5	18.8	85.5	424	17	BH670801	BH670801 BOMLK25TF
c	6	18.8	85.5	449	17	AZ009207	AZ009207 RPCI-23-3
c	7	18.8	85.5	476	10	BB768795	BB768795 BB768795
c	8	18.8	85.5	505	12	BF004802	BF004802 EST433300
	9	18.8	85.5	528	9	AA447445	AA447445 zw93g12.r
	10	18.8	85.5	544	17	AQ414714	AQ414714 RPCI-11-1
c	11	18.8	85.5	549	17	BH282218	BH282218 CH230-39A
	12	18.8	85.5	567	17	AQ534621	AQ534621 RPCI-11-3
	13	18.8	85.5	568	17	AZ506877	AZ506877 1M0348C14
	14	18.8	85.5	597	17	AQ534617	AQ534617 RPCI-11-3
c	15	18.8	85.5	710	17	BH552202	BH552202 BOHQN17TF
c	16	18.8	85.5	738	17	AZ572374	AZ572374 305PvD02
c	17	18.8	85.5	767	17	CNS07CKT	AL439299 T3 end of
	18	18.8	85.5	1059	17	CNS06H8C	AL398674 T3 end of
	19	18.8	85.5	1330	12	BF164624	BF164624 601772405
	20	18.4	83.6	479	17	CNS01OU1	AL153977 Anopheles
c	21	18	81.8	507	10	AW829796	AW829796 ra44e07.y
c	22	18	81.8	582	10	AW828026	AW828026 ra60g03.y
c	23	18	81.8	597	10	AW871607	AW871607 ra79h11.y
c	24	18	81.8	624	10	AW870709	AW870709 ra49f02.y
c	25	18	81.8	651	10	AW735590	AW735590 ra25b05.y
c	26	17.8	80.9	229	10	AV380205	AV380205 AV380205
c	27	17.8	80.9	280	10	BB381647	BB381647 BB381647
c	28	17.8	80.9	283	10	BB106733	BB106733 BB106733
c	29	17.8	80.9	318	9	AL837465	AL837465 AL837465
	30	17.8	80.9	318	13	BI034283	BI034283 PM2-NN016
	31	17.8	80.9	501	10	AV873095	AV873095 AV873095
c	32	17.8	80.9	510	17	AZ035704	AZ035704 RPCI-23-3
	33	17.8	80.9	527	10	AV873962	AV873962 AV873962
	34	17.8	80.9	533	14	BQ234625	BQ234625 hd47h10.g
c	35	17.8	80.9	539	17	BH427923	BH427923 BOGEA10TR
c	36	17.8	80.9	558	10	AW034365	AW034365 EST277936
	37	17.8	80.9	632	17	AZ417128	AZ417128 1M0192J12
	38	17.8	80.9	636	10	AV872452	AV872452 AV872452
	39	17.8	80.9	658	17	BH059762	BH059762 RPCI-24-2
c	40	17.8	80.9	666	10	BB641601	BB641601 BB641601
	41	17.8	80.9	666	17	BH717507	BH717507 BOHVE40TF
	42	17.8	80.9	670	17	DR10N24S	AL748390 Danio rer
	43	17.8	80.9	703	10	AV886791	AV886791 AV886791
	44	17.8	80.9	775	17	BH705808	BH705808 BOMPU04TR
	45	17.8	80.9	781	17	BH059760	BH059760 RPCI-24-2

ALIGNMENTS

RESULT 1
AW740409
LOCUS AW740409 342 bp mRNA linear EST 27-APR-2000
DEFINITION BRI10552 Biomphalaria glabrata (BS-90)-unexposed Lambda Zap Library Biomphalaria glabrata cDNA clone RBGIH55TR, mRNA sequence.
ACCESSION AW740409
VERSION AW740409.1 GI:7651688
KEYWORDS EST.
SOURCE bloodfluke planorb.
ORGANISM Biomphalaria glabrata
Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora ; Planorbidae; Biomphalaria.
REFERENCE 1 (bases 1 to 342)
AUTHORS Raghavan,N., Miller,A., Gardner,M., Kerlavage,A.R., FitzGerald,P.C. , Lewis,F.A. and Knight,M.
TITLE Genes expressed by the hemocytes of Biomphalaria glabrata before and after exposure to miracidia
JOURNAL Unpublished (2000)
COMMENT Contact: Raghavan N
Biomedical Research Institute
12111 Parklawn Dr., Rockville, MD 20852, USA
Tel: 301-881-3300 ext.128
Fax: 301-770-4756
Email: nkr@helix.nih.gov, snailsrule@aol.com.
FEATURES Location/Qualifiers
source 1. .342
/organism="Biomphalaria glabrata"
/strain="BS-90"
/db_xref="taxon:6526"

/clone="RBGIH55TR"
/clone_lib="Biomphalaria glabrata (BS-90)-unexposed Lambda Zap Library"
/sex="hermaphrodite"
/cell_type="Hemocyte"
/lab_host="Laboratory host"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Total RNA was isolated from the hemocytes of unexposed Biomphalaria glabrata (BS-90) snails and first strand cDNA synthesized using an oligo-dT primer-linker (XhoI). Second strand synthesis was followed by the ligation of EcoRI adaptors. Following digestion with XhoI, the completed, directional cDNA was cloned into Uni-ZAP XR Phagemid vector by Stratgene."

BASE COUNT 132 a 68 c 62 g 80 t
ORIGIN

Query Match 92.7%; Score 20.4; DB 10; Length 342;
Best Local Similarity 95.5%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAGCAA 22
||||||||||||||||||| ||
Db 174 AAGAAAAAATCTAGACAAGAAA 195

RESULT 2
BG360474
LOCUS BG360474 342 bp mRNA linear EST 07-MAR-2001
DEFINITION BRI10646 Biomphalaria glabrata (BS-90)-unexposed Lambda Zap Library Biomphalaria glabrata cDNA clone RBGIJ88TR, mRNA sequence.
ACCESSION BG360474
VERSION BG360474.1 GI:13243488
KEYWORDS EST.
SOURCE bloodfluke planorb.
ORGANISM Biomphalaria glabrata
Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora; Planorbidae; Biomphalaria.
REFERENCE 1 (bases 1 to 342)
AUTHORS Raghavan,N., Miller,A., Gardner,M., Kerlavage,A.R., FitzGerald,P.C., Lewis,F.A. and Knight,M.
TITLE Genes expressed by the hemocytes of Biomphalaria glabrata before and after exposure to miracidia
JOURNAL Unpublished (2000)
COMMENT Contact: Raghavan N
Biomedical Research Institute
12111 Parklawn Dr., Rockville, MD 20852, USA
Tel: 301-881-3300 ext.128
Fax: 301-770-4756
Email: nkr@helix.nih.gov, snailsrule@aol.com.
FEATURES Location/Qualifiers
source 1. .342
/organism="Biomphalaria glabrata"
/strain="BS-90"
/db_xref="taxon:6526"
/clone="RBGIJ88TR"
/clone_lib="Biomphalaria glabrata (BS-90)-unexposed Lambda Zap Library"
/sex="hermaphrodite"
/cell_type="Hemocyte"
/lab_host="Laboratory host"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Total RNA was isolated from the hemocytes of unexposed Biomphalaria glabrata (BS-90) snails and first strand cDNA synthesized using an oligo-dT primer-linker (XhoI). Second strand synthesis was followed by the ligation of EcoRI adaptors. Following digestion with XhoI, the completed, directional cDNA was cloned into Uni-ZAP XR Phagemid vector by Stratgene."
BASE COUNT 133 a 68 c 63 g 78 t
ORIGIN

Query Match 92.7%; Score 20.4; DB 12; Length 342;
Best Local Similarity 95.5%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAGCAA 22
||||||||||||||||||| ||
Db 174 AAGAAAAAATCTAGACAAGAAA 195

RESULT 3
AW740410
LOCUS AW740410 360 bp mRNA linear EST 27-APR-2000
DEFINITION BRI10553 Biomphalaria glabrata (BS-90)-unexposed Lambda Zap Library Biomphalaria glabrata cDNA clone RBGIH56TR, mRNA sequence.
ACCESSION AW740410
VERSION AW740410.1 GI:7651689
KEYWORDS EST.
SOURCE bloodfluke planorb.
ORGANISM Biomphalaria glabrata
Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora; Planorbidae; Biomphalaria.
REFERENCE 1 (bases 1 to 360)
AUTHORS Raghavan,N., Miller,A., Gardner,M., Kerlavage,A.R., FitzGerald,P.C., Lewis,F.A. and Knight,M.
TITLE Genes expressed by the hemocytes of Biomphalaria glabrata before and after exposure to miracidia
JOURNAL Unpublished (2000)
COMMENT Contact: Raghavan N
Biomedical Research Institute
12111 Parklawn Dr., Rockville, MD 20852, USA
Tel: 301-881-3300 ext.128
Fax: 301-770-4756
Email: nkr@helix.nih.gov, snailsrule@aol.com.
FEATURES Location/Qualifiers
source 1. .360
/organism="Biomphalaria glabrata"
/strain="BS-90"
/db_xref="taxon:6526"
/clone="RBGIH56TR"
/clone_lib="Biomphalaria glabrata (BS-90)-unexposed Lambda Zap Library"
/sex="hermaphrodite"
/cell_type="Hemocyte"
/lab_host="Laboratory host"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Total RNA was isolated from the hemocytes of unexposed Biomphalaria glabrata (BS-90) snails and first strand cDNA synthesized using an oligo-dT primer-linker (XhoI). Second strand synthesis was followed by the ligation of EcoRI adaptors. Following digestion with XhoI, the completed, directional cDNA was cloned into Uni-ZAP XR Phagemid vector by Stratgene."
BASE COUNT 137 a 72 c 62 g 86 t 3 others
ORIGIN

Query Match 92.7%; Score 20.4; DB 10; Length 360;
Best Local Similarity 95.5%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGAAAAAATCTAGACAAGCAA 22
||||||||||||||||||| ||
Db 174 AAGAAAAAATCTAGACAAGAAA 195

RESULT 4
BB016617
LOCUS BB016617 224 bp mRNA linear EST 22-JUN-2000
DEFINITION BB016617 RIKEN full-length enriched, adult male testis (DH10B) Mus musculus cDNA clone 4930563D02 3', mRNA sequence.
ACCESSION BB016617
VERSION BB016617.1 GI:8187765
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 224)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .224
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4930563D02"
/clone_lib="RIKEN full-length enriched, adult male testis (DH10B)"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT 99 a 28 c 39 g 58 t
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 224;
Best Local Similarity 90.9%; Pred. No. 4.2e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           ||||||||||  ||||||||||
Db     160 AAGAAAAAATTCAGACAAGCAA 181
```

RESULT 5
BH670801
LOCUS BH670801 424 bp DNA linear GSS 19-FEB-2002
DEFINITION BOMLK25TF BO_2_3_KB Brassica oleracea genomic clone BOMLK25, DNA sequence.
ACCESSION BH670801
VERSION BH670801.1 GI:18734592
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 424)
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .424
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOMLK25"
/clone_lib="BO_2_3_KB"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 165 a 86 c 66 g 107 t
ORIGIN

Query Match 85.5%; Score 18.8; DB 17; Length 424;
Best Local Similarity 90.9%; Pred. No. 3.8e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           ||||  ||||||||||  ||||||
Db     127 AAGATAAAATCTAGATAAGCAA 148
```

RESULT 6
AZ009207/c
LOCUS AZ009207 449 bp DNA linear GSS 25-FEB-2000
DEFINITION RPCI-23-365C18.TV RPCI-23 Mus musculus genomic clone RPCI-23-365C18 , DNA sequence.
ACCESSION AZ009207
VERSION AZ009207.1 GI:7084591
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 449)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)

```
COMMENT     Other_GSSs: RPCI-23-365C18.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 365  row: C  column: 18
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-365C18"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT       97 a     79 c     80 g    193 t
ORIGIN

  Query Match             85.5%;  Score 18.8;  DB 17;  Length 449;
  Best Local Similarity   90.9%;  Pred. No. 3.8e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           || |||||||||||||| |||||
Db     423 AATAAAAAATCTAGACTAGCAA 402


RESULT 7
BB768795/c
LOCUS       BB768795                476 bp    mRNA    linear   EST 17-OCT-2001
DEFINITION  BB768795 RIKEN full-length enriched, B16 F10Y cells Mus musculus
            cDNA clone G370089D24 3', mRNA sequence.
ACCESSION   BB768795
VERSION     BB768795.1  GI:16211337
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .476
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="G370089D24"
                     /clone_lib="RIKEN full-length enriched, B16 F10Y cells"
                     /cell_type="B16 F10Y cells"
                     /note="pooled tissues ; (tissue_type=cerebellum,
                     dev_stage=16 days neonate, sex=mixed),
                     (tissue_type=cerebellum, dev_stage=0 day neonate,
                     sex=mixed), (tissue_type=hippocampus, dev_stage=adult,
                     sex=male), (tissue_type=whole body, dev_stage=9 days
                     embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days
                     embryo, sex=mixed)"
BASE COUNT      128 a    105 c    111 g    132 t
ORIGIN

  Query Match             85.5%;  Score 18.8;  DB 10;  Length 476;
  Best Local Similarity   90.9%;  Pred. No. 3.8e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           || |||||| ||||||||||||
Db     303 AAAAAAAAACCTAGACAAGCAA 282


RESULT 8
BF004802/c
LOCUS       BF004802                505 bp    mRNA    linear   EST 06-OCT-2000
DEFINITION  EST433300 KV1 Medicago truncatula cDNA clone pKV1-18B14, mRNA
            sequence.
ACCESSION   BF004802
VERSION     BF004802.1  GI:10705077
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 505)
  AUTHORS   VandenBosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L.,
            Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and
            Fraser,C.M.
  TITLE     ESTs from roots of Medicago truncatula 24 hours after inoculation
            with Sinorhizobium meliloti
  JOURNAL   Unpublished (1999)
```

```
COMMENT     Contact: VandenBosch K
            Department of Plant Biology
            University of Minnesota
            220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
            Tel: 612 624 2755
            Fax: 612 625 1738
            Email: kvandenb@cbs.umn.edu
            Texas A&M University name: T268612e TIGR sequence name: MTIBJ07TK
            More information is available at: http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .505
                     /organism="Medicago truncatula"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pKV1-18B14"
                     /clone_lib="KV1"
                     /tissue_type="Seedling roots"
                     /dev_stage="24 hours post-inoculation with Sinorhizobium
                     meliloti"
                     /lab_host="E. coli strain XLOLR"
                     /note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                     was directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts.  Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in XLOLR cells."
BASE COUNT      144 a     93 c    103 g    164 t      1 others
ORIGIN

  Query Match            85.5%;  Score 18.8;  DB 12;  Length 505;
  Best Local Similarity  90.9%;  Pred. No. 3.7e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 AAGAAAAAATCTAGACAAGCAA 22
          ||||||| |||| |||||||||
Db    109 AAGAAAACATCTGGACAAGCAA 88


RESULT 9
AA447445
LOCUS       AA447445                528 bp    mRNA    linear   EST 04-JUN-1997
DEFINITION  zw93g12.r1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
            IMAGE:784582 5', mRNA sequence.
ACCESSION   AA447445
VERSION     AA447445.1  GI:2161115
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.
            , Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
            ,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 475.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Homo sapiens"
                     /db_xref="GDB:5982200"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:784582"
                     /clone_lib="Soares_total_fetus_Nb2HF8_9w"
                     /dev_stage="8-9 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from pooled 8-9 week
                     (total) fetus material with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      155 a    113 c     91 g    169 t
ORIGIN

  Query Match            85.5%;  Score 18.8;  DB 9;  Length 528;
  Best Local Similarity  90.9%;  Pred. No. 3.7e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 AAGAAAAAATCTAGACAAGCAA 22
          ||||||| |||| |||||||||
Db     24 AAGAAAACATCTGGACAAGCAA 45


RESULT 10
AQ414714
LOCUS       AQ414714                544 bp    DNA     linear   GSS 23-MAR-1999
DEFINITION  RPCI-11-171O19.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-171O19
            , DNA sequence.
ACCESSION   AQ414714
VERSION     AQ414714.1  GI:4473683
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Homo sapiens"
                     /db_xref="GDB:7565634"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-171O19"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      189 a     97 c    100 g    158 t
ORIGIN
```

```
  Query Match             85.5%;  Score 18.8;  DB 17;  Length 544;
  Best Local Similarity   90.9%;  Pred. No. 3.7e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           ||||||||| |||||||||  ||
Db     488 AAGAAAAAAGCTAGACAAGGAA 509


RESULT 11
BH282218/c
LOCUS       BH282218                549 bp    DNA     linear   GSS 30-NOV-2001
DEFINITION  CH230-39A11.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-39A11, DNA sequence.
ACCESSION   BH282218
VERSION     BH282218.1  GI:17194620
KEYWORDS    GSS.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-39A11.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 39  row: A  column: 11
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Rattus norvegicus"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-39A11"
                     /clone_lib="CHORI-230 Segment 1"
                     /sex="Female"
                     /cell_type="Brain"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      118 a    106 c    125 g    200 t
ORIGIN

  Query Match             85.5%;  Score 18.8;  DB 17;  Length 549;
  Best Local Similarity   90.9%;  Pred. No. 3.7e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           ||||||||| |||||||| |||
Db      43 AAGAAAAAAACTAGACAAACAA 22


RESULT 12
AQ534621
LOCUS       AQ534621                567 bp    DNA     linear   GSS 18-MAY-1999
DEFINITION  RPCI-11-353P23.TV RPCI-11 Homo sapiens genomic clone RPCI-11-353P23
            , DNA sequence.
ACCESSION   AQ534621
VERSION     AQ534621.1  GI:4846311
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-353P23.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Homo sapiens"
                     /db_xref="GDB:7635550"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-353P23"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      191 a     99 c    111 g    165 t      1 others
ORIGIN

  Query Match             85.5%;  Score 18.8;  DB 17;  Length 567;
  Best Local Similarity   90.9%;  Pred. No. 3.7e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           ||||||||| |||||||||  ||
Db     476 AAGAAAAAAGCTAGACAAGGAA 497


RESULT 13
AZ506877
LOCUS       AZ506877                568 bp    DNA     linear   GSS 05-OCT-2000
DEFINITION  1M0348C14F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0348C14 F, DNA sequence.
ACCESSION   AZ506877
VERSION     AZ506877.1  GI:10688193
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
```

```
  JOURNAL    Unpublished (2000)
  COMMENT    Contact: Robert B. Weiss
             University of Utah Genome Center
             University of Utah
             Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
             84112, USA
             Tel: 801 585 5606
             Fax: 801 585 7177
             Email: ddunn@genetics.utah.edu
             Insert Length: 10000   Std Error: 0.00
             Plate: 0348  row: C  column: 14
             Seq primer: CGTTGTAAAACGACGGCCAGT
             Class: plasmid ends
             High quality sequence stop: 568.
  FEATURES             Location/Qualifiers
       source          1. .568
                       /organism="Mus musculus"
                       /strain="C57BL/6J"
                       /db_xref="taxon:10090"
                       /clone="UUGC1M0348C14"
                       /clone_lib="Mouse 10kb plasmid UUGC1M library"
                       /sex="Male"
                       /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                       /note="Vector: PWD42nv; Purified genomic DNA from M.
                       musculus C57BL/6J (male) was obtained  from the Jackson
                       Laboratory Mouse DNA Resource
                       (http://www.jax.org/resources/documents/dnares/). The DNA
                       was hydrodynamically sheared by repeated passage through a
                       0.005 inch orifice at constant velocity. The sheared DNA
                       was blunt end-repaired with T4 DNA polymerase and T4
                       polynucleotide kinase. Adaptor oligonucleotides were
                       ligated to the blunt ends in high  molar excess. The
                       adaptored DNA was purified and size-selected for a 9.5 to
                       10.5 kb range using preparative agarose gel
                       electrophoresis. Vector DNA was prepared from a derivative
                       of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                       inducible derivative of plasmid R1. The vector was ligated
                       with adaptors complementary to the insert adaptors and
                       purified. The sheared, adaptored mouse DNA was annealed to
                       adaptored vector DNA, and transformed into
                       chemically-competent E. coli XL10-Gold (Stratagene) cells
                       and selected for ampicillin resistance."
  BASE COUNT      192 a    101 c    123 g    150 t      2 others
  ORIGIN

    Query Match             85.5%;  Score 18.8;  DB 17;  Length 568;
    Best Local Similarity   90.9%;  Pred. No. 3.7e+03;
    Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

  Qy       1 AAGAAAAAATCTAGACAAGCAA 22
             ||||||| |||||||||| ||
  Db     191 AAGAAAAGATCTAGACAAGAAA 212


  RESULT 14
  AQ534617
  LOCUS       AQ534617                597 bp    DNA     linear   GSS 18-MAY-1999
  DEFINITION  RPCI-11-353P21.TV RPCI-11 Homo sapiens genomic clone RPCI-11-353P21
              , DNA sequence.
  ACCESSION   AQ534617
  VERSION     AQ534617.1  GI:4846307
  KEYWORDS    GSS.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 597)
    AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
              ,J.C.
    TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
              Map Building
    JOURNAL   Unpublished (1997)
  COMMENT     Other_GSSs: RPCI-11-353P21.TJ
              Contact: Shaying Zhao, William Nierman, Mark Adams
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: hbe@tigr.org
              Clones are derived from the human BAC library RPCI-11. For BAC
              library availability, please contact Pieter de Jong
              (pieter@dejong.med.buffalo.edu). Clones may be purchased from
              BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
              Research Genet cs (info@resgen.com). BAC end search page:
              http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
              Seq primer: T7
              Class: BAC ends.
  FEATURES             Location/Qualifiers
       source          1. .597
                       /organism="Homo sapiens"
                       /db_xref="GDB:7635548"
                       /db_xref="taxon:9606"
                       /clone="RPCI-11-353P21"
                       /clone_lib="RPCI-11"
                       /sex="Male"
                       /cell_type="Lymphocytes"
                       /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                       RPCI11 Human Male BAC Library"
  BASE COUNT      201 a    107 c    115 g    174 t
  ORIGIN

    Query Match             85.5%;  Score 18.8;  DB 17;  Length 597;
    Best Local Similarity   90.9%;  Pred. No. 3.6e+03;
    Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

  Qy       1 AAGAAAAAATCTAGACAAGCAA 22
             ||||||||| ||||||||| ||
  Db     496 AAGAAAAAAGCTAGACAAGGAA 517


  RESULT 15
  BH552202/c
  LOCUS       BH552202                710 bp    DNA     linear   GSS 14-DEC-2001
  DEFINITION  BOHQN17TF BOHQ Brassica oleracea genomic clone BOHQN17, DNA
              sequence.
  ACCESSION   BH552202
  VERSION     BH552202.1  GI:17803982
  KEYWORDS    GSS.
  SOURCE      Brassica oleracea.
    ORGANISM  Brassica oleracea
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
  REFERENCE   1  (bases 1 to 710)
    AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
    TITLE     Whole genome shotgun sequencing of Brassica oleracea
    JOURNAL   Unpublished (2001)
  COMMENT     Other_GSSs: BOHQN17TR
              Contact: Chris Town
              TIGR
              9712 Medical Center Drive, Rockville, MD 20850, USA.
              Tel: 301-838-3523
              Fax: 301-838-0208
              Email: cdtown@tigr.org
              DNA is from a doubled haploid provided by Tom Osborn.
              Seq primer: TF
              Class: sheared ends.
  FEATURES             Location/Qualifiers
       source          1. .710
                       /organism="Brassica oleracea"
                       /strain="TO1000DH3"
                       /db_xref="taxon:3712"
                       /clone="BOHQN17"
                       /clone_lib="BOHQ"
                       /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
```

genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 209 a 146 c 149 g 206 t
ORIGIN

Query Match 85.5%; Score 18.8; DB 17; Length 710;
Best Local Similarity 90.9%; Pred. No. 3.5e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 AAGAAAAAATCTAGACAAGCAA 22
           |||| |||||||||| ||||||
Db     154 AAGATAAAATCTAGATAAGCAA 133
```

Search completed: March 17, 2003, 13:09:20
Job time : 782.688 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:43:42 ; Search time 230.108 Seconds
(without alignments)
3161.870 Million cell updates/sec

Title: US-09-836-439-5
Perfect score: 25
Sequence: 1 gctttctttgctgagagctcttcca 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23.4	93.6	269	10	MUSOPS4	M36698 Mouse opsin
2	23.4	93.6	2610	10	BC031766	BC031766 Mus muscu
3	23.4	93.6	3249	10	BC013125	BC013125 Mus muscu
4	23.4	93.6	9483	10	MUSOPS	M55171 Mouse opsin
5	21.8	87.2	175	5	AF249155	AF249155 Indirana
6	21.8	87.2	1062	5	TNI293018	AJ293018 Tetraodon
c 7	20.4	81.6	178639	2	AC095437	AC095437 Rattus no
8	20.2	80.8	169	5	AF249131	AF249131 Microhyla
9	20.2	80.8	175	5	AF249129	AF249129 Bufo mela
10	20.2	80.8	175	5	AF249133	AF249133 Mantella
11	20.2	80.8	175	5	AF249134	AF249134 Mantidact
12	20.2	80.8	175	5	AF249136	AF249136 Boophis x
13	20.2	80.8	175	5	AF249137	AF249137 Boophis t
14	20.2	80.8	175	5	AF249138	AF249138 Laliostom
15	20.2	80.8	175	5	AF249139	AF249139 Fejervary
16	20.2	80.8	175	5	AF249140	AF249140 Fejervary
17	20.2	80.8	175	5	AF249141	AF249141 Hoplobatr
18	20.2	80.8	175	5	AF249142	AF249142 Sphaeroth
19	20.2	80.8	175	5	AF249143	AF249143 Euphlycti
20	20.2	80.8	175	5	AF249144	AF249144 Nannophry
21	20.2	80.8	175	5	AF249146	AF249146 Nyctibatr
22	20.2	80.8	175	5	AF249147	AF249147 Limnonect
23	20.2	80.8	175	5	AF249148	AF249148 Limnonect
24	20.2	80.8	175	5	AF249149	AF249149 Rana curt
25	20.2	80.8	175	5	AF249150	AF249150 Rana temp
26	20.2	80.8	175	5	AF249151	AF249151 Rana temp
27	20.2	80.8	175	5	AF249152	AF249152 Micrixalu
28	20.2	80.8	175	5	AF249154	AF249154 Indirana
29	20.2	80.8	175	5	AF249156	AF249156 Polypedat
30	20.2	80.8	175	5	AF249157	AF249157 Rhacophor
31	20.2	80.8	175	5	AF249158	AF249158 Philautus
32	20.2	80.8	405	5	AF221974	AF221974 Bufo pusi
33	20.2	80.8	416	5	AF221979	AF221979 Bufo pant
34	20.2	80.8	825	5	AF137213	AF137213 Ostracion
35	20.2	80.8	924	5	AB084931	AB084931 Dimidioch
36	20.2	80.8	924	5	AB084938	AB084938 Oreochrom
37	20.2	80.8	924	5	AB084940	AB084940 Sarothero
38	20.2	80.8	924	5	AB084941	AB084941 Spathodus
39	20.2	80.8	924	5	AB084944	AB084944 Tilapia r
40	20.2	80.8	924	5	AB084947	AB084947 Xenotilap
41	20.2	80.8	1053	5	AF021242	AF021242 Melopsitt
42	20.2	80.8	1059	5	AMU12328	U12328 Astyanax me
43	20.2	80.8	1399	5	RTU59920	U59920 Rana tempor
44	20.2	80.8	1423	5	MSUY18666	Y18666 Mullus surm
45	20.2	80.8	1469	5	MVU57539	U57539 Myripristis

ALIGNMENTS

RESULT 1
MUSOPS4
LOCUS MUSOPS4 269 bp DNA linear ROD 08-MAY-1993
DEFINITION Mouse opsin gene, exon 4.
ACCESSION M36698 X69175
VERSION M36698.1 GI:200149
KEYWORDS opsin.
SEGMENT 4 of 5
SOURCE Mus musculus (strain C56BL/6J) eye DNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 269)
AUTHORS Baehr,W., Falk,J.D., Bugra,K., Triantafyllos,J.T. and McGinnis,J.F.
TITLE Isolation and analysis of the mouse opsin gene

```
  JOURNAL   FEBS Lett. 238 (2), 253-256 (1988)
  MEDLINE   89005694
   PUBMED   2844600
FEATURES             Location/Qualifiers
     source          1. .269
                     /organism="Mus musculus"
                     /strain="C56BL/6J"
                     /db_xref="taxon:10090"
                     /tissue_type="eye"
     intron          order(M36697.1:170. .>189,<1. .14)
                     /gene="opsin"
                     /note="117 bp gap"
                     /number=3
     exon            15. .254
                     /gene="opsin"
                     /number=4
BASE COUNT       57 a     87 c     58 g     67 t
ORIGIN      About 117 bp after segment 3.

  Query Match             93.6%;  Score 23.4;  DB 10;  Length 269;
  Best Local Similarity   96.0%;  Pred. No. 1.4;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
           ||||||||||||| |||||||||||
Db     192 GCTTTCTTTGCTAAGAGCTCTTCCA 216


RESULT 2
BC031766
LOCUS       BC031766               2610 bp    mRNA    linear   ROD 07-AUG-2002
DEFINITION  Mus musculus, clone MGC:25387 IMAGE:4527040, mRNA, complete cds.
ACCESSION   BC031766
VERSION     BC031766.1  GI:21594394
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2610)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 31 Row: n Column: 2
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, GenomeScan gene prediction.
FEATURES             Location/Qualifiers
     source          1. .2610
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="MGC:25387 IMAGE:4527040"
                     /tissue_type="Eye, retina, mouse strain C57Bl\6"
                     /clone_lib="NIH_MGC_94"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             80. .1126
                     /codon_start=1
                     /product="Unknown (protein for MGC:25387)"
                     /protein_id="AAH31766.1"
                     /db_xref="GI:21594395"
                     /db_xref="LocusID:212541"
                     /translation="MNGTEGPNFYVPFSNVTGVVRSPFEQPQYYLAEPWQFSMLAAYM
                     FLLIVLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGY
                     FVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVVFT
                     WIMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPMIV
                     IFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIFFLICWLPYASVAFYIF
                     THQGSNFGPIFMTLPAFFAKSSSIYNPVIYIMLNKQFRNCMLTTLCCGKNPLGDDDAS
                     ATASKTETSQVAPA"
BASE COUNT      604 a    727 c    656 g    623 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 10;  Length 2610;
  Best Local Similarity   96.0%;  Pred. No. 1.5;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
           ||||||||||||| |||||||||||
Db     953 GCTTTCTTTGCTAAGAGCTCTTCCA 977


RESULT 3
BC013125
LOCUS       BC013125               3249 bp    mRNA    linear   ROD 07-AUG-2002
DEFINITION  Mus musculus, Similar to rhodopsin (opsin 2, rod pigment)
            (retinitis pigmentosa 4, autosomal dominant), clone MGC:21585
            IMAGE:4500760, mRNA, complete cds.
ACCESSION   BC013125
VERSION     BC013125.1  GI:15341884
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3249)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 28 Row: e Column: 21
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1. .3249
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
```

```
                     /clone="MGC:21585 IMAGE:4500760"
                     /tissue_type="Eye, retina, mouse strain C57Bl\6"
                     /clone_lib="NIH_MGC_94"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             79. .1125
                     /codon_start=1
                     /product="Similar to rhodopsin (opsin 2, rod pigment)
                     (retinitis pigmentosa 4, autosomal dominant)"
                     /protein_id="AAH13125.1"
                     /db_xref="GI:15341885"
                     /translation="MNGTEGPNFYVPFSNVTGVVRSPFEQPQYYLAEPWQFSMLAAYM
                     FLLIVLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGY
                     FVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVVFT
                     WIMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPMIV
                     IFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIFFLICWLPYASVAFYIF
                     THQGSNFGPIFMTLPAFFAKSSSIYNPVIYIMLNKQFRNCMLTTLCCGKNPLGDDDAS
                     ATASKTETSQVAPA"
BASE COUNT      766 a    867 c    829 g    787 t
ORIGIN

  Query Match           93.6%;  Score 23.4;  DB 10;  Length 3249;
  Best Local Similarity  96.0%;  Pred. No. 1.5;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||| ||||||||||||
Db    952 GCTTTCTTTGCTAAGAGCTCTTCCA 976


RESULT 4
MUSOPS
LOCUS       MUSOPS                  9483 bp    DNA     linear   ROD 07-MAR-1995
DEFINITION  Mouse opsin (MOPS) gene, complete cds.
ACCESSION   M55171
VERSION     M55171.1  GI:200144
KEYWORDS    opsin.
SOURCE      Mouse adult liver DNA, clone lambda-MO1.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 9483)
  AUTHORS   al-Ubaidi,M.R., Pittler,S.J., Champagne,M.S., Triantafyllos,J.T.,
            McGinnis,J.F. and Baehr,W.
  TITLE     Mouse opsin. Gene structure and molecular basis of multiple
            transcripts
  JOURNAL   J. Biol. Chem. 265 (33), 20563-20569 (1990)
  MEDLINE   91056108
   PUBMED   1978723
FEATURES             Location/Qualifiers
     source          1. .9483
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="chromosome 6"
                     /clone="lambda-MO1"
                     /tissue_type="liver"
                     /dev_stage="adult"
                     /tissue_lib="EMBL3"
     gene            join(1338. .1794,3283. .3451,4525. .4690,4808. .5047,
                     6019. .9483)
                     /gene="MOPS"
     mRNA            join(1338. .1794,3283. .3451,4525. .4690,4808. .5047,
                     6019. .9483)
                     /gene="MOPS"
                     /product="opsin"
     exon            1338. .1794
                     /gene="MOPS"
                     /number=1
     CDS             join(1434. .1794,3283. .3451,4525. .4690,4808. .5047,
                     6019. .6129)
                     /gene="MOPS"
                     /codon_start=1
                     /product="opsin"
                     /protein_id="AAA63392.1"
                     /db_xref="GI:200145"
                     /translation="MNGTEGPNFYVPFSNVTGVGRSPFEQPQYYLAEPWQFSMLAAYM
                     FLLIVLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGY
                     FVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVVFT
                     WIMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPMIV
                     IFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIFFLICWLPYASVAFYIF
                     THQGSNFGPIFMTLPAFFAKSSSIYNPVIYIMLNKQFRNCMLTTLCCGKNPLGDDDAS
                     ATASKTETSQVAPA"
     intron          1795. .3282
                     /gene="MOPS"
     exon            3283. .3451
                     /gene="MOPS"
                     /number=2
     intron          3452. .4524
                     /gene="MOPS"
     exon            4525. .4690
                     /gene="MOPS"
                     /number=3
     intron          4691. .4807
                     /gene="MOPS"
                     /number=3
     exon            4808. .5047
                     /gene="MOPS"
                     /number=4
     intron          5048. .6018
                     /gene="MOPS"
                     /number=4
     exon            6019. .9483
                     /gene="MOPS"
                     /number=5
     polyA_signal    6477. .6482
                     /gene="MOPS"
     polyA_signal    6862. .6877
                     /gene="MOPS"
     polyA_signal    7022. .7027
                     /gene="MOPS"
     polyA_signal    7565. .7570
                     /gene="MOPS"
     polyA_signal    8205. .8210
                     /gene="MOPS"
     polyA_signal    8900. .8905
                     /gene="MOPS"
     polyA_signal    9092. .9097
                     /gene="MOPS"
     polyA_signal    9303. .9309
                     /gene="MOPS"
     polyA_signal    9410. .9415
                     /gene="MOPS"
BASE COUNT     2236 a   2510 c   2444 g   2293 t
ORIGIN

  Query Match           93.6%;  Score 23.4;  DB 10;  Length 9483;
  Best Local Similarity  96.0%;  Pred. No. 1.5;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||| ||||||||||||
Db   4985 GCTTTCTTTGCTAAGAGCTCTTCCA 5009


RESULT 5
AF249155
LOCUS       AF249155                175 bp    DNA     linear   VRT 17-JAN-2001
DEFINITION  Indirana sp. 2 rhodopsin gene, exon 4 and partial cds.
ACCESSION   AF249155
VERSION     AF249155.1  GI:12247239
KEYWORDS    .
SOURCE      Indirana sp. 2.
  ORGANISM  Indirana sp. 2
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae;
            Indirana.
```

```
REFERENCE   1  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
            Reveal Co-variation between Larval and Adult Traits
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
            Molecular Biology and Medicine, rue Jeener and Brachet 12,
            Gosselies B-6041, Belgium
FEATURES             Location/Qualifiers
     source          1. .175
                     /organism="Indirana sp. 2"
                     /db_xref="taxon:147866"
     mRNA            <1. .>175
                     /product="rhodopsin"
     CDS             <1. .>175
                     /codon_start=2
                     /product="rhodopsin"
                     /protein_id="AAG49798.1"
                     /db_xref="GI:12247240"
                     /translation="AEKEVTRMVVIMVVFFLICWVPYAYVAFYIFTHQGSEFGPIFMT
                     VPAFFAKSSSIYNP"
     exon            <1. .>175
                     /number=4
BASE COUNT       34 a     54 c     37 g     49 t      1 others
ORIGIN

  Query Match             87.2%;  Score 21.8;  DB 5;  Length 175;
  Best Local Similarity   92.0%;  Pred. No. 8.6;
  Matches   23;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||| ||||||||||||
Db    140 GCTTTCTTTGCCAAGAGCTCTTCCA 164


RESULT 6
TNI293018
LOCUS       TNI293018               1062 bp    DNA     linear   VRT 23-AUG-2000
DEFINITION  Tetraodon nigroviridis rhod gene for rhodopsin.
ACCESSION   AJ293018
VERSION     AJ293018.1  GI:9909983
KEYWORDS    RHOD gene; rhodopsin.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1062)
  AUTHORS   Fischer,C.
  TITLE     Tetraodon nigroviridis gene for rhodopsin
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1062)
  AUTHORS   Fischer,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2000) Fischer C., Centre National de Sequencage,
            Genoscope, 2 rue Gaston Cremieux, CP5706, F-91057 Evry Cedex,
            FRANCE
FEATURES             Location/Qualifiers
     source          1. .1062
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
     gene            1. .1062
                     /gene="rhod"
     CDS             1. .1062
                     /gene="rhod"
                     /codon_start=1
                     /product="rhodopsin"
                     /protein_id="CAC04526.1"
                     /db_xref="GI:9909984"
                     /translation="MNGTEGPFFYVPMLNTTGIVRSPYDYPQYYLVNPAAYAALGAYM
                     FLLILLGFPINFLTLYVTIEHKKLRTPLNYILLNLAVANLFMVFGGFTTTMYTSMHGY
                     FVLGRLGCNLEGFFATLGGEIGLWSLVVLAIERWMVVCKPISNFRFGENHAIMGLAFT
                     WIMACACAVPPLVGWSRYIPEGMQCSCGVDYYTRAEGFNNESFVVYMFICHFLIPMAV
                     VFFCYGRLLCAVKEAAAAQQESETTQRAEREVTRMVVIMVVAFLICWLPYAGVAWWIF
                     THQGSEFGPVFMTIPAFFAKSSSIYNPLIYICMNKQFRHCMITTLCCGKNPFEEEEGA
                     STTSKTEASSVSSSSVSPA"
BASE COUNT      198 a    329 c    259 g    276 t
ORIGIN

  Query Match             87.2%;  Score 21.8;  DB 5;  Length 1062;
  Best Local Similarity   92.0%;  Pred. No. 8.9;
  Matches   23;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          || |||||||| ||||||||||||
Db    874 GCCTTCTTTGCTAAGAGCTCTTCCA 898


RESULT 7
AC095437/c
LOCUS       AC095437               178639 bp    DNA     linear   HTG 10-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-4H9, *** SEQUENCING IN PROGRESS ***,
            67 unordered pieces.
ACCESSION   AC095437
VERSION     AC095437.3  GI:21716847
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 178639)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
```

```
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178639)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 178639)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 9, 2002 this sequence version replaced gi:17941914.
            ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: GCHA
                Center clone name: CH230-4H9
            ---------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 128571 bases at least Q40
                Consensus quality: 133786 bases at least Q30
                Consensus quality: 137554 bases at least Q20
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 67 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1036: contig of 1036 bp in length
            *     1037     1136: gap of unknown length
            *     1137     2377: contig of 1241 bp in length
            *     2378     2477: gap of unknown length
            *     2478     3634: contig of 1157 bp in length
            *     3635     3734: gap of unknown length
            *     3735     5027: contig of 1293 bp in length
            *     5028     5127: gap of unknown length
            *     5128     6327: contig of 1200 bp in length
            *     6328     6427: gap of unknown length
            *     6428     8034: contig of 1607 bp in length
            *     8035     8134: gap of unknown length
            *     8135     9424: contig of 1290 bp in length
            *     9425     9524: gap of unknown length
            *     9525    10938: contig of 1414 bp in length
            *    10939    11038: gap of unknown length
            *    11039    12077: contig of 1039 bp in length
            *    12078    12177: gap of unknown length
            *    12178    13190: contig of 1013 bp in length
            *    13191    13290: gap of unknown length
            *    13291    14394: contig of 1104 bp in length
            *    14395    14494: gap of unknown length
            *    14495    15666: contig of 1172 bp in length
            *    15667    15766: gap of unknown length
            *    15767    17029: contig of 1263 bp in length
            *    17030    17129: gap of unknown length
            *    17130    18244: contig of 1115 bp in length
            *    18245    18344: gap of unknown length
            *    18345    19987: contig of 1643 bp in length
            *    19988    20087: gap of unknown length
            *    20088    21161: contig of 1074 bp in length
            *    21162    21261: gap of unknown length
            *    21262    22935: contig of 1674 bp in length
            *    22936    23035: gap of unknown length
            *    23036    24726: contig of 1691 bp in length
            *    24727    24826: gap of unknown length
            *    24827    25984: contig of 1158 bp in length
            *    25985    26084: gap of unknown length
            *    26085    28117: contig of 2033 bp in length
            *    28118    28217: gap of unknown length
            *    28218    29226: contig of 1009 bp in length
            *    29227    29326: gap of unknown length
            *    29327    31330: contig of 2004 bp in length
            *    31331    31430: gap of unknown length
            *    31431    33008: contig of 1578 bp in length
            *    33009    33108: gap of unknown length
            *    33109    34893: contig of 1785 bp in length
            *    34894    34993: gap of unknown length
            *    34994    36418: contig of 1425 bp in length
            *    36419    36518: gap of unknown length
            *    36519    38755: contig of 2237 bp in length
            *    38756    38855: gap of unknown length
            *    38856    40715: contig of 1860 bp in length
            *    40716    40815: gap of unknown length
            *    40816    42591: contig of 1776 bp in length
            *    42592    42691: gap of unknown length
            *    42692    44432: contig of 1741 bp in length
            *    44433    44532: gap of unknown length
            *    44533    46670: contig of 2138 bp in length
            *    46671    46770: gap of unknown length
            *    46771    48171: contig of 1401 bp in length
            *    48172    48271: gap of unknown length
            *    48272    50293: contig of 2022 bp in length
            *    50294    50393: gap of unknown length
            *    50394    52099: contig of 1706 bp in length
            *    52100    52199: gap of unknown length
            *    52200    54742: contig of 2543 bp in length
            *    54743    54842: gap of unknown length
            *    54843    57839: contig of 2997 bp in length
            *    57840    57939: gap of unknown length
            *    57940    59989: contig of 2050 bp in length
            *    59990    60089: gap of unknown length
            *    60090    61599: contig of 1510 bp in length
            *    61600    61699: gap of unknown length
            *    61700    63235: contig of 1536 bp in length
            *    63236    63335: gap of unknown length
            *    63336    65696: contig of 2361 bp in length
            *    65697    65796: gap of unknown length
            *    65797    68230: contig of 2434 bp in length
            *    68231    68330: gap of unknown length
            *    68331    70088: contig of 1758 bp in length
            *    70089    70188: gap of unknown length
            *    70189    72866: contig of 2678 bp in length
            *    72867    72966: gap of unknown length
            *    72967    76485: contig of 3519 bp in length
            *    76486    76585: gap of unknown length
            *    76586    78904: contig of 2319 bp in length
            *    78905    79004: gap of unknown length
            *    79005    81763: contig of 2759 bp in length
            *    81764    81863: gap of unknown length
            *    81864    85381: contig of 3518 bp in length
            *    85382    85481: gap of unknown length
            *    85482    88489: contig of 3008 bp in length
            *    88490    88589: gap of unknown length
            *    88590    91317: contig of 2728 bp in length
            *    91318    91417: gap of unknown length
            *    91418    93443: contig of 2026 bp in length
            *    93444    93543: gap of unknown length
            *    93544    95582: contig of 2039 bp in length
            *    95583    95682: gap of unknown length
            *    95683    99140: contig of 3458 bp in length
            *    99141    99240: gap of unknown length
            *    99241   101817: contig of 2577 bp in length
            *   101818   101917: gap of unknown length
            *   101918   105395: contig of 3478 bp in length
```

```
          *   105396   105495: gap of unknown length
          *   105496   110101: contig of 4606 bp in length

  Query Match             81.6%;  Score 20.4;  DB 2;  Length 178639;
  Best Local Similarity   95.5%;  Pred. No. 47;
  Matches   21;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 CTTTCTTTGCTGAGAGCTCTTC 23
          ||||||||||||| ||||||||
Db 164970 CTTTCTTTGCTGAAAGCTCTTC 164949


RESULT 8
AF249131
LOCUS       AF249131                169 bp    DNA     linear   VRT 17-JAN-2001
DEFINITION  Microhyla sp. rhodopsin gene, exon 4 and partial cds.
ACCESSION   AF249131
VERSION     AF249131.1  GI:12247191
KEYWORDS    .
SOURCE      Microhyla sp.
  ORGANISM  Microhyla sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Microhyloidea;
            Microhylidae; Microhylinae; Microhyla.
REFERENCE   1  (bases 1 to 169)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
            Reveal Co-variation between Larval and Adult Traits
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
            Molecular Biology and Medicine, rue Jeener and Brachet 12,
            Gosselies B-6041, Belgium
FEATURES             Location/Qualifiers
     source          1. .169
                     /organism="Microhyla sp."
                     /db_xref="taxon:147864"
     mRNA            <1. .>169
                     /product="rhodopsin"
     CDS             <1. .>169
                     /codon_start=2
                     /product="rhodopsin"
                     /protein_id="AAG49774.1"
                     /db_xref="GI:12247192"
                     /translation="AEKEVTRMVVIMVIFFLICWVPYAYVAFYIFTHQGSEFGPIFMT
                     VPAFFAKSSSIY"
     exon            <1. .>169
                     /number=4
BASE COUNT       32 a     46 c     34 g     57 t
ORIGIN

  Query Match             80.8%;  Score 20.2;  DB 5;  Length 169;
  Best Local Similarity   88.0%;  Pred. No. 52;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          || ||||||||  ||||||||||||
Db    140 GCCTTCTTTGCCAAGAGCTCTTCCA 164


RESULT 9
AF249129
LOCUS       AF249129                175 bp    DNA     linear   VRT 17-JAN-2001
DEFINITION  Bufo melanostictus rhodopsin gene, exon 4 and partial cds.
ACCESSION   AF249129
VERSION     AF249129.1  GI:12247187
KEYWORDS    .
SOURCE      Bufo melanostictus.
  ORGANISM  Bufo melanostictus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae;
            Bufo.
REFERENCE   1  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
            Reveal Co-variation between Larval and Adult Traits
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
            Molecular Biology and Medicine, rue Jeener and Brachet 12,
            Gosselies B-6041, Belgium
FEATURES             Location/Qualifiers
     source          1. .175
                     /organism="Bufo melanostictus"
                     /db_xref="taxon:30335"
     mRNA            <1. .>175
                     /product="rhodopsin"
     CDS             <1. .>175
                     /codon_start=2
                     /product="rhodopsin"
                     /protein_id="AAG49772.1"
                     /db_xref="GI:12247188"
                     /translation="AEKEVTRMVIIMVVFFLICWVPYATVAFFIFTHQGSEXGPIXMT
                     IPAFFAKSSSIYNP"
     exon            <1. .>175
                     /number=4
BASE COUNT       31 a     54 c     37 g     50 t      3 others
ORIGIN

  Query Match             80.8%;  Score 20.2;  DB 5;  Length 175;
  Best Local Similarity   88.0%;  Pred. No. 52;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||  |||| |||||||
Db    140 GCTTTCTTTGCCAAGAGTTCTTCCA 164


RESULT 10
AF249133
LOCUS       AF249133                175 bp    DNA     linear   VRT 17-JAN-2001
DEFINITION  Mantella madagascariensis rhodopsin gene, exon 4 and partial cds.
ACCESSION   AF249133
VERSION     AF249133.1  GI:12247195
KEYWORDS    .
SOURCE      Mantella madagascariensis.
  ORGANISM  Mantella madagascariensis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Rhacophoridae;
            Mantella.
REFERENCE   1  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
            Reveal Co-variation between Larval and Adult Traits
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
            Molecular Biology and Medicine, rue Jeener and Brachet 12,
            Gosselies B-6041, Belgium
FEATURES             Location/Qualifiers
     source          1. .175
                     /organism="Mantella madagascariensis"
                     /db_xref="taxon:111290"
     mRNA            <1. .>175
                     /product="rhodopsin"
     CDS             <1. .>175
                     /codon_start=2
                     /product="rhodopsin"
                     /protein_id="AAG49776.1"
                     /db_xref="GI:12247196"
```

```
                     /translation="AEKEVTRMVVIMVVFFLICWVPYAYVAFYIFTHQGSEFGPIFMT
                     VPAFFAKSSAIYNP"
     exon            <1. .>175
                     /number=4
BASE COUNT       33 a     53 c     36 g     52 t      1 others
ORIGIN

  Query Match             80.8%;  Score 20.2;  DB 5;  Length 175;
  Best Local Similarity   88.0%;  Pred. No. 52;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          |||||||||||  ||||||||  |||
Db    140 GCTTTCTTTGCCAAGAGCTCTGCCA 164


RESULT 11
AF249134
LOCUS       AF249134                175 bp    DNA     linear   VRT 17-JAN-2001
DEFINITION  Mantidactylus cf. ulcerosus rhodopsin gene, exon 4 and partial cds.
ACCESSION   AF249134
VERSION     AF249134.1  GI:12247197
KEYWORDS    .
SOURCE      Mantidactylus cf. ulcerosus.
  ORGANISM  Mantidactylus cf. ulcerosus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Rhacophoridae;
            Mantidactylus.
REFERENCE   1  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
            Reveal Co-variation between Larval and Adult Traits
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
            Molecular Biology and Medicine, rue Jeener and Brachet 12,
            Gosselies B-6041, Belgium
FEATURES             Location/Qualifiers
     source          1. .175
                     /organism="Mantidactylus cf. ulcerosus"
                     /db_xref="taxon:129014"
     mRNA            <1. .>175
                     /product="rhodopsin"
     CDS             <1. .>175
                     /codon_start=2
                     /product="rhodopsin"
                     /protein_id="AAG49777.1"
                     /db_xref="GI:12247198"
                     /translation="AEKEVTRMVVIMVVFFLICWVPYAYVAFYIFTHQGSEFGPIFMT
                     VPAFFAKSSAIYNP"
     exon            <1. .>175
                     /number=4
BASE COUNT       36 a     50 c     36 g     53 t
ORIGIN

  Query Match             80.8%;  Score 20.2;  DB 5;  Length 175;
  Best Local Similarity   88.0%;  Pred. No. 52;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          |||||||||||  ||||||||  |||
Db    140 GCTTTCTTTGCCAAGAGCTCTGCCA 164


RESULT 12
AF249136
LOCUS       AF249136                175 bp    DNA     linear   VRT 17-JAN-2001
DEFINITION  Boophis xerophilus rhodopsin gene, exon 4 and partial cds.
ACCESSION   AF249136
VERSION     AF249136.1  GI:12247201
KEYWORDS    .
SOURCE      Boophis xerophilus.
  ORGANISM  Boophis xerophilus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Rhacophoridae;
            Boophis.
REFERENCE   1  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
            Reveal Co-variation between Larval and Adult Traits
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
            Molecular Biology and Medicine, rue Jeener and Brachet 12,
            Gosselies B-6041, Belgium
FEATURES             Location/Qualifiers
     source          1. .175
                     /organism="Boophis xerophilus"
                     /db_xref="taxon:128996"
     mRNA            <1. .>175
                     /product="rhodopsin"
     CDS             <1. .>175
                     /codon_start=2
                     /product="rhodopsin"
                     /protein_id="AAG49779.1"
                     /db_xref="GI:12247202"
                     /translation="AEKEVTRMVIIMVVFFLICWVPYAYVAFYIFTHQGSEFGPIFMT
                     VPAFFAKSSAIYNP"
     exon            <1. .>175
                     /number=4
BASE COUNT       34 a     55 c     34 g     52 t
ORIGIN

  Query Match             80.8%;  Score 20.2;  DB 5;  Length 175;
  Best Local Similarity   88.0%;  Pred. No. 52;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          |||||||||||  ||||||||  |||
Db    140 GCTTTCTTTGCCAAGAGCTCTGCCA 164


RESULT 13
AF249137
LOCUS       AF249137                175 bp    DNA     linear   VRT 17-JAN-2001
DEFINITION  Boophis tephraeomystax rhodopsin gene, exon 4 and partial cds.
ACCESSION   AF249137
VERSION     AF249137.1  GI:12247203
KEYWORDS    .
SOURCE      Boophis tephraeomystax.
  ORGANISM  Boophis tephraeomystax
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Rhacophoridae;
            Boophis.
REFERENCE   1  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
            Reveal Co-variation between Larval and Adult Traits
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
            Molecular Biology and Medicine, rue Jeener and Brachet 12,
            Gosselies B-6041, Belgium
FEATURES             Location/Qualifiers
     source          1. .175
                     /organism="Boophis tephraeomystax"
                     /db_xref="taxon:68440"
     mRNA            <1. .>175
                     /product="rhodopsin"
     CDS             <1. .>175
```

```
                     /codon_start=2
                     /product="rhodopsin"
                     /protein_id="AAG49780.1"
                     /db_xref="GI:12247204"
                     /translation="AEKEVTRMVIIMVIFFLICWVPYAYVAFYIFTHQGSEFGPIFMT
                     VPAFFAKSSAIYNP"
     exon            <1. .>175
                     /number=4
BASE COUNT       35 a     56 c     33 g     51 t
ORIGIN

  Query Match             80.8%;  Score 20.2;  DB 5;  Length 175;
  Best Local Similarity   88.0%;  Pred. No. 52;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||| |||||||| |||
Db    140 GCTTTCTTTGCCAAGAGCTCTGCCA 164


RESULT 14
AF249138
LOCUS       AF249138                 175 bp    DNA     linear   VRT 17-JAN-2001
DEFINITION  Laliostoma labrosum rhodopsin gene, exon 4 and partial cds.
ACCESSION   AF249138
VERSION     AF249138.1  GI:12247205
KEYWORDS    .
SOURCE      Laliostoma labrosum.
  ORGANISM  Laliostoma labrosum
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae;
            Laliostoma.
REFERENCE   1  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
            Reveal Co-variation between Larval and Adult Traits
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
            Molecular Biology and Medicine, rue Jeener and Brachet 12,
            Gosselies B-6041, Belgium
FEATURES             Location/Qualifiers
     source          1. .175
                     /organism="Laliostoma labrosum"
                     /db_xref="taxon:129011"
     mRNA            <1. .>175
                     /product="rhodopsin"
     CDS             <1. .>175
                     /codon_start=2
                     /product="rhodopsin"
                     /protein_id="AAG49781.1"
                     /db_xref="GI:12247206"
                     /translation="AEKEVTRMVIIMVVFFLICWVPYAYVAFYIFTHQGSEFGPIFMT
                     VPAFFAKSSAIYNP"
     exon            <1. .>175
                     /number=4
BASE COUNT       35 a     55 c     36 g     49 t
ORIGIN

  Query Match             80.8%;  Score 20.2;  DB 5;  Length 175;
  Best Local Similarity   88.0%;  Pred. No. 52;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||| |||||||| |||
Db    140 GCTTTCTTTGCCAAGAGCTCTGCCA 164


RESULT 15
AF249139
LOCUS       AF249139                 175 bp    DNA     linear   VRT 17-JAN-2001
DEFINITION  Fejervarya syhadrensis rhodopsin gene, exon 4 and partial cds.
ACCESSION   AF249139
VERSION     AF249139.1  GI:12247207
KEYWORDS    .
SOURCE      Fejervarya syhadrensis.
  ORGANISM  Fejervarya syhadrensis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae;
            Fejervarya.
REFERENCE   1  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
            Reveal Co-variation between Larval and Adult Traits
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175)
  AUTHORS   Bossuyt,F. and Milinkovitch,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
            Molecular Biology and Medicine, rue Jeener and Brachet 12,
            Gosselies B-6041, Belgium
FEATURES             Location/Qualifiers
     source          1. .175
                     /organism="Fejervarya syhadrensis"
                     /db_xref="taxon:128998"
     mRNA            <1. .>175
                     /product="rhodopsin"
     CDS             <1. .>175
                     /codon_start=2
                     /product="rhodopsin"
                     /protein_id="AAG49782.1"
                     /db_xref="GI:12247208"
                     /translation="AEKEVTRMVIIMVVFFCICWFPYAFVAFYIFTHQGSEFGPVFMT
                     VPAFFAKSSAIYNP"
     exon            <1. .>175
                     /number=4
BASE COUNT       34 a     50 c     35 g     54 t      2 others
ORIGIN

  Query Match             80.8%;  Score 20.2;  DB 5;  Length 175;
  Best Local Similarity   88.0%;  Pred. No. 52;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||| |||||||| |||
Db    140 GCTTTCTTTGCCAAGAGCTCTGCCA 164


Search completed: March 17, 2003, 11:31:57
Job time : 256.108 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:42:12 ; Search time 143.28 Seconds
(without alignments)
392.938 Million cell updates/sec

Title: US-09-836-439-5
Perfect score: 25
Sequence: 1 gctttctttgctgagagctcttcca 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	18.2	72.8	807	21	AAC48879	Arabidopsis thalia
c	2	18.2	72.8	906	23	AAS77337	DNA encoding novel
	3	18.2	72.8	3254	22	AAC85295	Mouse SPANK cDNA.
	4	18.2	72.8	6912	22	AAS36854	Human cardiovascul
	5	17.8	71.2	702	23	ABL26659	Drosophila melanog
c	6	17.8	71.2	756	21	AAH51547	Human MGSTII relat
c	7	17.8	71.2	1000	21	AAH51581	Human MGSTII relat
c	8	17.8	71.2	1000	22	AAS30820	Human cDNA encodin
	9	17.8	71.2	2573	19	AAV59153	Human ReproSA-1 co
	10	17.8	71.2	3153	23	ABL26658	Drosophila melanog
	11	17.8	71.2	3870	23	ABL26632	Drosophila melanog
	12	17.8	71.2	8002	24	AAD31629	Arabidopsis thalia
c	13	17.6	70.4	223	21	AAC07949	Human secreted pro
	14	17.6	70.4	342	24	ABL69358	Prostate cancer re
c	15	17.6	70.4	384	23	ABV11069	Human prostate exp
	16	17.6	70.4	403	24	ABN26150	Human ORFX polynuc
c	17	17.6	70.4	437	23	ABV32215	Human prostate exp
c	18	17.6	70.4	437	23	ABV41146	Human prostate exp
c	19	17.6	70.4	437	23	ABV41150	Human prostate exp
c	20	17.6	70.4	474	24	ABA92961	Human cDNA clone C
c	21	17.6	70.4	745	20	AAZ41246	Human normal ovari
c	22	17.6	70.4	749	23	AAS88970	DNA encoding novel
c	23	17.6	70.4	761	20	AAZ42214	Human normal bladd
c	24	17.6	70.4	779	21	AAC78068	Human cancer assoc
c	25	17.6	70.4	5252	22	AAF30010	Rat CARD-6 cDNA.
c	26	17.6	70.4	23899	23	ABL10362	Drosophila melanog
	27	17.4	69.6	2849	24	ABL53927	Human dihydroorota
	28	17.2	68.8	570	22	AAL13764	Human breast cance
c	29	17.2	68.8	687	21	AAF13683	Aspergillus oryzae
	30	17.2	68.8	712	22	AAF22438	Human breast cance
	31	17.2	68.8	712	22	AAF22514	Human breast cance
	32	17.2	68.8	811	22	AAL22631	Human breast cance
c	33	17.2	68.8	850	21	AAA64643	Partial sequence M
c	34	17.2	68.8	904	21	AAA64645	Partial sequence M
	35	17.2	68.8	1512	22	AAF30257	Human cDNA encodin
	36	17.2	68.8	1578	22	AAI59927	Human polynucleoti
	37	17.2	68.8	1676	21	AAC75873	Human ORFX ORF1428
	38	17.2	68.8	1739	22	AAK94496	Human full-length
	39	17.2	68.8	1828	22	AAH15348	Human cDNA sequenc
	40	17.2	68.8	1934	22	AAD12593	Human protein havi
	41	17.2	68.8	1945	22	AAD04905	Human secreted pro
c	42	17.2	68.8	1945	24	ABL90687	Human polynucleoti
	43	17.2	68.8	2015	22	AAC87005	Nucleotide sequenc
	44	17.2	68.8	2015	22	AAA91020	Human secreted pro
	45	17.2	68.8	2015	22	AAF44267	Human PRO7170 nucl

ALIGNMENTS

RESULT 1
AAC48879/c
ID AAC48879 standard; DNA; 807 BP.
XX
AC AAC48879;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 59111.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

```
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             72.8%;  Score 18.2;  DB 21;  Length 807;
  Best Local Similarity   87.0%;  Pred. No. 1.4e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 CTTTCTTTGCTGAGAGCTCTTCC 24
           |||  |||||||||| |||||||
Db     401 CTTGTTTTGCTGAGATCTCTTCC 379


RESULT 2
AAS77337/c
ID   AAS77337 standard; cDNA; 906 BP.
XX
AC   AAS77337;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #13141.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG13150.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 13141; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 906 BP; 243 A; 218 C; 269 G; 176 T; 0 other;

  Query Match             72.8%;  Score 18.2;  DB 23;  Length 906;
  Best Local Similarity   87.0%;  Pred. No. 1.4e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCTGAGAGCTCTTC 23
           |||||||||||| ||  |||||
Db     677 GCTTTCTTTGCTGCGATGTCTTC 655


RESULT 3
AAC85295
ID   AAC85295 standard; cDNA; 3254 BP.
XX
AC   AAC85295;
XX
DT   29-MAR-2001  (first entry)
XX
DE   Mouse SPANK cDNA.
XX
KW   SPANK; SAM; sterile alpha motif; PARP; insulin resistance;
KW   poly adenosine diphosphate-ribose polymerase; catalytic domain;
KW   ANK; ankyrin repeat; cytosol; insulin-responsive aminopeptidase;
KW   IRAP; GLUT4; adipocyte; insulin signalling pathway; hyperlipidaemia;
KW   glucose intolerance; atheromatous disease; atherosclerosis;
KW   obesity; cardiac insufficiency; coronary insufficiency; stroke;
KW   high blood pressure; non-insulin dependent diabetes; hypertension;
KW   hyperuricemia; Syndrome X; muscular dystrophy; muscle atrophy; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             308..3254
FT                   /*tag= a
FT                   /product= "Murine SPANK"
FT                   /transl_except= (pos:1754,aa:Gly)
FT                   /note= "This codon has an apparent 2 nucleotide
FT                         deletion, which alters the reading frame"
FT                   /transl_except= (pos:2145,aa:DTQIGKLLIG)
FT                   /note= "The nucleotides encoding residues 614 to 623
FT                         are absent, which alters the reading frame"
FT                   /transl_except= (pos:2602..2606,aa:Met)
FT                   /note= "This codon has an apparent 2 nucleotide
FT                         insertion, which alters the reading frame"
XX
PN   WO200077225-A1.
XX
PD   21-DEC-2000.
XX
PF   09-JUN-2000; 2000WO-US15926.
XX
```

PR 11-JUN-1999; 99US-0138957.
XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Chi N, Lodish HF;
XX
DR WPI; 2001-091404/10.
DR P-PSDB; AAB47023.
XX
PT New insulin signalling protein SPANK, useful for reducing body mass,
PT glucose intolerance or insulin resistance and for preventing or
PT treating obesity-related and muscle-related diseases -
XX
PS Claim 3; Fig 5A; 65pp; English.
XX
CC This sequence represents the mouse SPANK cDNA. The SPANK
CC protein comprises 3 domains:
CC (a) a SAM (sterile alpha motif) domain;
CC (b) a PARP (poly adenosine diphosphate-ribose polymerase) catalytic
CC domain; and
CC (c) an ANK domain composed of ankyrin repeats.
CC SPANK is a cytosolic protein which can poly(ADP-ribosyl)ate itself.
CC SPANK binds insulin-responsive aminopeptidase (IRAP) and modulates
CC translocation of GLUT4 in the perinuclear region of adipocytes. It
CC is an effector in the insulin signalling pathway in eukaryotic cells.
CC SPANK is useful for reducing body mass, reducing glucose
CC intolerance or insulin resistance, for preventing or treating
CC obesity-related diseases or disorders, such as obesity, cardiac
CC insufficiency, coronary insufficiency, stroke, hypertension,
CC atheromatous disease, atherosclerosis, high blood pressure, non-insulin
CC dependent diabetes, hyperlipidaemia, hyperuricemia and Syndrome X and is
CC also useful for preventing or treating muscle-related diseases or
CC disorders, such as muscular dystrophy, muscle atrophy and muscle
CC fatigue. Antibodies immunospecific for SPANK are useful for detecting
CC the presence of SPANK polypeptide in a biological sample.
XX
SQ Sequence 3254 BP; 871 A; 746 C; 863 G; 774 T; 0 other;

Query Match 72.8%; Score 18.2; DB 22; Length 3254;
Best Local Similarity 87.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 TTTCTTTGCTGAGAGCTCTTCCA 25
|| |||||||||| |||||||
Db 2930 TTATTTTGCTGAGAACTCTTCCA 2952

RESULT 4
AAS36854
ID AAS36854 standard; DNA; 6912 BP.
XX
AC AAS36854;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human cardiovascular system antigen genomic DNA SEQ ID No 2354.
XX
KW Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; tissue regeneration;
KW anti-infertility.
XX
OS Homo sapiens.
XX
PN WO200155321-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01340.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX

DR WPI; 2001-451930/48.
XX
PT New cardiovascular system related polynucleotides and polypeptides,
PT useful for diagnosing, treating and/or preventing disorders of the
PT cardiovascular system -
XX
PS Claim 1; SEQ ID No 2354; 674pp; English.
XX
CC Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
CC the cardiovascular system antigen polypeptides of the invention.
CC Cardiovascular system antigens and their associated polynucleotides are
CC useful in the diagnosis, treatment and prevention of various types of
CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. A pathological condition can be determined by
CC detecting the presence or absence of a mutation in a cardiovascular
CC system antigen polynucleotide. The treatable disorders include autoimmune
CC diseases such as rheumatoid arthritis, hyperproliferative disorders such
CC as neoplasms of the breast or liver, cardiovascular disorders such as
CC cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
CC nervous system disorders such as Alzheimer's disease, infections caused
CC by bacteria, viruses and fungi, ocular disorders such as corneal
CC infection, endocrine disorders such as premature labour and infertility,
CC gastrointestinal disorders such as Crohn's disease, renal disorders such
CC as glomerulonephritis and respiratory disorders such as asthma and
CC pleurisy. The polypeptides can also be used to aid wound healing, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, to regenerate tissues and in chemotaxis.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

Query Match 72.8%; Score 18.2; DB 22; Length 6912;
Best Local Similarity 87.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 TTTCTTTGCTGAGAGCTCTTCCA 25
||||||||||||| ||| |||||
Db 647 TTTCTTTGCTGATAGCGGTTCCA 669

RESULT 5
ABL26659
ID ABL26659 standard; DNA; 702 BP.
XX
AC ABL26659;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 31450.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -
XX
PS Claim 1; SEQ ID NO 31450; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 702 BP; 152 A; 190 C; 158 G; 202 T; 0 other;

Query Match 71.2%; Score 17.8; DB 23; Length 702;
Best Local Similarity 90.5%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 TCTTTGCTGAGAGCTCTTCCA 25
||||| ||||| |||||||||
Db 475 TCTTTCCTGAGCGCTCTTCCA 495

RESULT 6
AAH51547/c
ID AAH51547 standard; DNA; 756 BP.
XX
AC AAH51547;
XX
DT 29-AUG-2001 (first entry)
XX
DE Human MGSTII related DNA containing a biallelic polymorphism SEQ ID 438.
XX
KW Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII;
KW microsomal glutathione S-transferase II; malate decarboxylase enzyme;
KW DME1; ME1; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5;
KW flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5;
KW dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype;
KW phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype;
KW uridine diphosphate glucoronosyl transferase; UGT2; asthma; hepatoxicity;
KW zileuton; ds.
XX
OS Homo sapiens.
XX
PN WO200058508-A2.
XX
PD 05-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-IB00403.
XX
PR 25-MAR-1999; 99US-0126269.
PR 30-APR-1999; 99US-0131961.
XX
PA (GEST) GENSET.
XX
PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX
DR WPI; 2000-638353/61.
XX
PT Polynucleotides comprising sequences from malate decarboxylase
PT enzyme-related biallelic markers used for genotyping -
XX
PS Claim 13; Page 616; 673pp; English.
XX
CC Sequences AAH51110-AAH51593 represent human DNA fragments which contain
CC biallelic markers. The sequences are related to various human genes
CC including microsomal glutathione S-transferase II (MGSTII), malate
CC decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione
CC reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),
CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate
CC dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine
CC diphosphate glucoronosyl transferases (UGT2). Each of these sequences
CC contains a biallelic marker/polymorphism, which is represented in the
CC sequence as a degenerate/undefined base. The genes to which the biallelic
CC marker containing sequences are related are involved in drug metabolism.
CC Sequences AAH51594 - AAH51598 represent the genomic sequence of the
CC MGSTII gene and four alternative MGSTII cDNA sequences. AAB62905-AAB62906
CC are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in
CC an example for the amplification of human genomic DNA fragments. The
CC invention includes a method of genotyping comprising determining the
CC identity of a nucleotide at a DME- or MGSTII-related biallelic marker in
CC a biological sample. The method is used to determine the frequency in
CC population of an allele of a DME- or MGST-II related biallelic marker and
CC to select an individual for inclusion in a clinical trial of a drug
CC treatment. The method is also used to detect association between allele
CC and phenotype, and to detect association between haplotype and phenotype.
CC The polynucleotides are used, in hybridization assays, sequencing assays
CC or allele specific amplification assays. The method can be used to
CC determine whether an individual suffers or is at risk of developing
CC asthma or is at risk of developing hepatoxicity on treatment with
CC zileuton.
XX
SQ Sequence 756 BP; 224 A; 140 C; 177 G; 213 T; 2 other;

Query Match 71.2%; Score 17.8; DB 21; Length 756;
Best Local Similarity 90.5%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTTCTTTGCTGAGAGCTCTTC 23
||||||||| |||||||| |
Db 614 TTTCTTTGCTCAGAGCTCTCC 594

RESULT 7
AAH51581/c
ID AAH51581 standard; DNA; 1000 BP.
XX
AC AAH51581;
XX
DT 29-AUG-2001 (first entry)
XX
DE Human MGSTII related DNA containing a biallelic polymorphism SEQ ID 472.
XX
KW Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII;
KW microsomal glutathione S-transferase II; malate decarboxylase enzyme;
KW DME1; ME1; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5;
KW flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5;
KW dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype;
KW phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype;
KW uridine diphosphate glucoronosyl transferase; UGT2; asthma; hepatoxicity;
KW zileuton; ds.
XX
OS Homo sapiens.
XX
PN WO200058508-A2.
XX
PD 05-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-IB00403.
XX
PR 25-MAR-1999; 99US-0126269.
PR 30-APR-1999; 99US-0131961.
XX
PA (GEST) GENSET.
XX
PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX
DR WPI; 2000-638353/61.
XX
PT Polynucleotides comprising sequences from malate decarboxylase
PT enzyme-related biallelic markers used for genotyping -
XX

```
PS   Claim 13; Page 641; 673pp; English.
XX
CC   Sequences AAH51110-AAH51593 represent human DNA fragments which contain
CC   biallelic markers. The sequences are related to various human genes
CC   including microsomal glutathione S-transferase II (MGSTII), malate
CC   decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione
CC   reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),
CC   gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate
CC   dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine
CC   diphosphate glucoronosyl transferases (UGT2). Each of these sequences
CC   contains a biallelic marker/polymorphism, which is represented in the
CC   sequence as a degenerate/undefined base. The genes to which the biallelic
CC   marker containing sequences are related are involved in drug metabolism.
CC   Sequences AAH51594 - AAH51598 represent the genomic sequence of the
CC   MGSTII gene and four alternative MGSTII cDNA sequences. AAB62905-AAB62906
CC   are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in
CC   an example for the amplification of human genomic DNA fragments. The
CC   invention includes a method of genotyping comprising determining the
CC   identity of a nucleotide at a DME- or MGSTII-related biallelic marker in
CC   a biological sample. The method is used to determine the frequency in
CC   population of an allele of a DME- or MGST-II related biallelic marker and
CC   to select an individual for inclusion in a clinical trial of a drug
CC   treatment. The method is also used to detect association between allele
CC   and phenotype, and to detect association between haplotype and phenotype.
CC   The polynucleotides are used, in hybridization assays, sequencing assays
CC   or allele specific amplification assays. The method can be used to
CC   determine whether an individual suffers or is at risk of developing
CC   asthma or is at risk of developing hepatoxicity on treatment with
CC   zileuton.
XX
SQ   Sequence 1000 BP; 300 A; 200 C; 211 G; 288 T; 1 other;

  Query Match             71.2%;  Score 17.8;  DB 21;  Length 1000;
  Best Local Similarity   90.5%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 TTTCTTTGCTGAGAGCTCTTC 23
           |||||||||| |||||||| |
Db     851 TTTCTTTGCTCAGAGCTCTCC 831


RESULT 8
AAS30820/c
ID   AAS30820 standard; cDNA; 1000 BP.
XX
AC   AAS30820;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Human cDNA encoding G protein-coupled receptor nGPCR-2428.
XX
KW   Human; G protein-coupled receptor; nGPCR-x; ss; antiviral; analgesic;
KW   cytostatic; cardiant; antidiabetic; anoretic; hypotensive; hypertensive;
KW   antiparkinsonian; nootropic; neuroprotective; antidepressant;
KW   viral infection; HIV-1; human immunodeficiency virus; HIV-2; pain;
KW   cancer; metabolic disease; cardiovascular disease; type 2 diabetes;
KW   obesity; anorexia; hypotension; hypertension; myocardial infarction;
KW   atherosclerosis; Parkinson's disease; psychosis; neurological disorder;
KW   schizophrenia; migraine; major depression; anxiety; mental disorder;
KW   manic depression; dyskinesia; Huntington's disease; Tourette's Syndrome.
XX
OS   Homo sapiens.
XX
PN   WO200166750-A2.
XX
PD   13-SEP-2001.
XX
PF   08-MAR-2001; 2001WO-US07322.
XX
PR   08-MAR-2000; 2000US-0187581.
PR   08-MAR-2000; 2000US-0187582.
PR   08-MAR-2000; 2000US-0187714.
PR   08-MAR-2000; 2000US-0187715.
PR   08-MAR-2000; 2000US-0187825.
PR   08-MAR-2000; 2000US-0187828.
PR   08-MAR-2000; 2000US-0187829.
PR   08-MAR-2000; 2000US-0187830.
PR   08-MAR-2000; 2000US-0187833.
PR   08-MAR-2000; 2000US-0187874.
PR   08-MAR-2000; 2000US-0187930.
PR   08-MAR-2000; 2000US-0188049.
PR   08-MAR-2000; 2000US-0189294.
PR   08-MAR-2000; 2000US-0187929.
PR   08-MAR-2000; 2000US-0187928.
XX
PA   (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI   Vogeli G,  Wood LS;
XX
DR   WPI; 2001-536778/59.
DR   P-PSDB; AAU19251.
XX
PT   Isolated nucleic acid molecules encoding G protein-coupled receptors
PT   termed nGPCR-x, useful in the treatment and diagnosis of viral
PT   infections, cancers and mental disorders (e.g. Parkinson's disease and
PT   schizophrenia) -
XX
PS   Claim 4; Page 221-222; 336pp; English.
XX
CC   The invention relates to novel isolated nucleic acid molecules encoding
CC   G protein-coupled receptors termed nGPCR-x. nGPCR-x polynucleotides,
CC   polypeptides, and modulators may be used in the treatment of diseases and
CC   conditions such as infections, such as viral infections caused by HIV-1
CC   (human immunodeficiency virus) or HIV-2, pain, cancers, metabolic and
CC   cardiovascular diseases and disorders (e.g., type 2 diabetes, obesity,
CC   anorexia, hypotension, hypertension, myocardial infarction,
CC   atherosclerosis), Parkinson's disease, and psychotic and
CC   neurological disorders, including schizophrenia, migraine, major
CC   depression, anxiety, mental disorder, manic depression, and
CC   dyskinesias, such as Huntington's disease or Tourette's Syndrome
CC   and many other diseases and syndromes listed in the specification.
CC   nGPCR-x polynucleotides and polypeptides, as well as nGPCR-x
CC   modulators, may also be used in diagnostic assays for such diseases or
CC   conditions. The present sequence encodes a G protein-coupled
CC   receptor of the invention.
XX
SQ   Sequence 1000 BP; 265 A; 208 C; 299 G; 228 T; 0 other;

  Query Match             71.2%;  Score 17.8;  DB 22;  Length 1000;
  Best Local Similarity   90.5%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 TTTCTTTGCTGAGAGCTCTTC 23
           |||||||||| |||||||| |
Db     791 TTTCTTTGCTCAGAGCTCTCC 771


RESULT 9
AAV59153
ID   AAV59153 standard; cDNA; 2573 BP.
XX
AC   AAV59153;
XX
DT   14-DEC-1998  (first entry)
XX
DE   Human ReproSA-1 coding sequence.
XX
KW   ss; human; ReproSA-1; contraceptive; anti-sperm; fertility;
KW   in vitro fertilisation.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             125..1654
FT                   /*tag= a
FT                   /product= "ReproSA-1 protein"
```

```
XX
PN   WO9836073-A1.
XX
PD   20-AUG-1998.
XX
PF   17-FEB-1998;   98WO-US03513.
XX
PR   18-FEB-1997;   97US-0039574.
XX
PA   (REPR-) REPROGEN INC.
XX
PI   French CK,  Neilson LI,  Schneider PA,  Yamamoto KK;
XX
DR   WPI; 1998-480771/41.
DR   P-PSDB; AAW77323.
XX
PT   Sperm-specific polypeptide ReproSA-1 and related nucleic acid -
PT   transformed cells and antibodies, useful in contraceptive vaccines
PT   or compositions, and for diagnosis and treatment of infertility
XX
PS   Claim 16; Fig 1; 62pp; English.
XX
CC   The ReproSA-1 polypeptide, it's fragments, or nucleic acid encoding them
CC   are used in contraceptive vaccines to generate an anti-sperm response.
CC   Antibodies can also be used as topical contraceptive (spermicide), also
CC   antisense sequences, ribozymes and triplex-forming molecules.  Detection,
CC   by reaction with immobilised peptides, of anti-(I) antibodies (in serum,
CC   semen, saliva, cervical or vaginal mucosa) is used to diagnose
CC   immunological infertility and to monitor the effect of vaccination.
CC   nucleotide fragments are used to diagnose infertility-associated
CC   mutations in the gene encoding Repro-SA, and for isolation of related
CC   sequences.  Peptides and anti-idiotypic antibodies against the
CC   antibodies are used to inactivate the antibodies i.e. to increase
CC   fertility and may be added to sperm samples before in vitro
CC   fertilisation.
XX
SQ   Sequence 2573 BP; 754 A; 532 C; 586 G; 701 T; 0 other;

  Query Match             71.2%;  Score 17.8;  DB 19;  Length 2573;
  Best Local Similarity   90.5%;  Pred. No. 2.5e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 TTTCTTTGCTGAGAGCTCTTC 23
           | ||||||||||||| ||||||
Db     252 TGTCTTTGCTGAGAACTCTTC 272


RESULT 10
ABL26658
ID   ABL26658 standard; DNA; 3153 BP.
XX
AC   ABL26658;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 31447.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 31447; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3153 BP; 864 A; 693 C; 665 G; 931 T; 0 other;

  Query Match             71.2%;  Score 17.8;  DB 23;  Length 3153;
  Best Local Similarity   90.5%;  Pred. No. 2.6e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 TCTTTGCTGAGAGCTCTTCCA 25
           ||||| ||||| |||||||||
Db    1746 TCTTTCCTGAGCGCTCTTCCA 1766


RESULT 11
ABL26632
ID   ABL26632 standard; DNA; 3870 BP.
XX
AC   ABL26632;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 31369.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 31369; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
```

```
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3870 BP; 961 A; 917 C; 886 G; 1106 T; 0 other;

  Query Match             71.2%;  Score 17.8;  DB 23;  Length 3870;
  Best Local Similarity   90.5%;  Pred. No. 2.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 TCTTTGCTGAGAGCTCTTCCA 25
           ||||| ||||| |||||||||
Db     103 TCTTTCCTGAGCGCTCTTCCA 123


RESULT 12
AAD31629
ID   AAD31629 standard; DNA; 8002 BP.
XX
AC   AAD31629;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Arabidopsis thaliana GT1209 gene.
XX
KW   GT1209 gene; herbicide; plant tissue; cell; seedling growth; development;
KW   ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200212273-A2.
XX
PD   14-FEB-2002.
XX
PF   01-AUG-2001; 2001WO-EP08910.
XX
PR   03-AUG-2000; 2000US-222779P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Levin JZ,  Wegrich Glover L,  Budziszewski GJ;
XX
DR   WPI; 2002-241730/29.
XX
PT   New polypeptide having GT1209, GT1354 or GT0946 activity, obtained from
PT   Arabidopsis, useful as herbicide targets in screening assays to
PT   identify the inhibitors or potential herbicides  -
XX
PS   Disclosure; Page 72-74; 82pp; English.
XX
CC   The invention relates to a polypeptide having GT1209, GT1354 or GT0946
CC   activity obtained from Arabidopsis. The invention may also be applied to
CC   the development of herbicide tolerant plants, plant tissues, plant seeds
CC   and plant cells. The polypeptide is useful as herbicide targets in
CC   screening assays to identify potential herbicides and inhibitors of
CC   GT1209, GT1354 or GT0946 activity. A compound having herbicidal activity
CC   is useful for suppressing the growth of a plant. The newly discovered
CC   GT1209, GT1354 or GT0946 genes are essential for seedling growth and
CC   development. The present sequence is Arabidopsis thaliana GT1209 gene.
XX
SQ   Sequence 8002 BP; 2119 A; 1489 C; 1502 G; 2892 T; 0 other;

  Query Match             71.2%;  Score 17.8;  DB 24;  Length 8002;
  Best Local Similarity   90.5%;  Pred. No. 3e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 CTTTCTTTGCTGAGAGCTCTT 22
           |||||||||||||  ||||||
Db    6994 CTTTCTTTGCTGATTGCTCTT 7014


RESULT 13
AAC07949/c
ID   AAC07949 standard; cDNA; 223 BP.
XX
AC   AAC07949;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 12024.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 12024; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 223 BP; 77 A; 38 C; 53 G; 54 T; 1 other;

  Query Match             70.4%;  Score 17.6;  DB 21;  Length 223;
  Best Local Similarity   83.3%;  Pred. No. 2e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCTGAGAGCTCTTCC 24
           || |||||||| ||||  ||||||
Db     152 GCCTTCTTTGCAGAGATTTCTTCC 129


RESULT 14
ABL69358
ID   ABL69358 standard; DNA; 342 BP.
XX
AC   ABL69358;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Prostate cancer related gene sequence SEQ ID NO:7695.
XX
```

KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 7695; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ Sequence 342 BP; 36 A; 109 C; 92 G; 105 T; 0 other;

Query Match 70.4%; Score 17.6; DB 24; Length 342;
Best Local Similarity 83.3%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCTTTCTTTGCTGAGAGCTCTTCC 24
||||||||| ||| | | ||||||
Db 67 GCTTTCTTTTCTGGGGGATCTTCC 90

RESULT 15
ABV11069/c
ID ABV11069 standard; cDNA; 384 BP.
XX
AC ABV11069;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 11060.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 1792; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer

CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 384 BP; 139 A; 87 C; 73 G; 85 T; 0 other;

Query Match 70.4%; Score 17.6; DB 23; Length 384;
Best Local Similarity 83.3%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy       2 CTTTCTTTGCTGAGAGCTCTTCCA 25
           ||||||||| ||||||| || ||
Db     374 CTTTCTTTGGGGAGAGCTATTTCA 351
```

Search completed: March 17, 2003, 10:50:50
Job time : 146.446 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 10:23:12 ; Search time 883.737 Seconds
(without alignments)
458.154 Million cell updates/sec

Title: US-09-836-439-5
Perfect score: 25
Sequence: 1 gctttctttgctgagagctcttcca 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	23.4	93.6	402	12	BE950666	BE950666 UI-M-CE0-
	2	23.4	93.6	495	13	BI737699	BI737699 603358627
	3	23.4	93.6	628	13	BI734160	BI734160 603351436
	4	23.4	93.6	636	13	BI736587	BI736587 603361053
	5	23.4	93.6	649	12	BG403992	BG403992 602419859
	6	23.4	93.6	651	13	BI730233	BI730233 603350201
	7	23.4	93.6	656	13	BI730180	BI730180 603349724
	8	23.4	93.6	663	13	BI729678	BI729678 603349362
	9	23.4	93.6	672	12	BG298553	BG298553 602396937
	10	23.4	93.6	676	13	BI736996	BI736996 603360842
	11	23.4	93.6	681	13	BI735044	BI735044 603356176
	12	23.4	93.6	682	12	BG295809	BG295809 602393292
	13	23.4	93.6	694	13	BI737008	BI737008 603360855
	14	23.4	93.6	696	12	BG298167	BG298167 602396304
	15	23.4	93.6	697	13	BI735675	BI735675 603357821
	16	23.4	93.6	697	14	BM937113	BM937113 UI-M-CE0-
	17	23.4	93.6	713	13	BI738044	BI738044 603357135
	18	23.4	93.6	717	13	BI734477	BI734477 603351934
	19	23.4	93.6	718	12	BG294123	BG294123 602390988
	20	23.4	93.6	720	13	BI730669	BI730669 603350310
	21	23.4	93.6	743	12	BG404627	BG404627 602420777
	22	23.4	93.6	744	13	BI873234	BI873234 603397523
	23	23.4	93.6	746	12	BG298242	BG298242 602396396
	24	23.4	93.6	759	12	BG293849	BG293849 602390360
	25	23.4	93.6	761	12	BG297332	BG297332 602395456
	26	23.4	93.6	773	13	BI872453	BI872453 603398026
	27	23.4	93.6	774	12	BG261526	BG261526 602373279
	28	23.4	93.6	783	13	BI736361	BI736361 603360075
	29	23.4	93.6	790	13	BI734568	BI734568 603356401
	30	23.4	93.6	802	13	BI732279	BI732279 603354207
	31	23.4	93.6	806	12	BG404226	BG404226 602420283
	32	23.4	93.6	825	13	BI729984	BI729984 603350079
	33	23.4	93.6	832	13	BI872123	BI872123 603396431
	34	23.4	93.6	837	12	BG293614	BG293614 602390274
	35	23.4	93.6	851	13	BI737918	BI737918 603357492
	36	23.4	93.6	854	13	BI739285	BI739285 603359578
	37	23.4	93.6	873	13	BI739561	BI739561 603361813
	38	23.4	93.6	879	13	BI737919	BI737919 603357493
	39	23.4	93.6	880	14	BQ932985	BQ932985 AGENCOURT
	40	23.4	93.6	903	12	BG293203	BG293203 602390159
	41	23.4	93.6	909	14	BQ921973	BQ921973 AGENCOURT
	42	23.4	93.6	914	14	BQ938546	BQ938546 AGENCOURT
	43	23.4	93.6	915	13	BI734393	BI734393 603351143
	44	23.4	93.6	917	12	BG262056	BG262056 602373890
	45	23.4	93.6	918	13	BI738365	BI738365 603361127

ALIGNMENTS

RESULT 1
BE950666/c
LOCUS BE950666 402 bp mRNA linear EST 29-APR-2002
DEFINITION UI-M-CE0-aza-d-06-0-UI.s1 NIH_BMAP_Ret3 Mus musculus cDNA clone UI-M-CE0-aza-d-06-0-UI 3', mRNA sequence.
ACCESSION BE950666
VERSION BE950666.1 GI:10589332
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 402)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site

```
            and the oligo-dT track served to verify it as a clone from the
            retina tissue cDNA Library Preparation: M.B. Soares Lab Clone
            distribution: Researchers may obtain BMAP cDNA clones from RESEARCH
            GENETICS. It should be noted that Bento Soares is generating a
            small number of additional specialized non-redundant arrays of BMAP
            cDNAs whose availability will be considered under appropriate and
            limited collaborative arrangements The tissue for this library was
            contributed by Dr. Xin-Yuan Fu, Yale University School of Medicine
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .402
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UI-M-CE0-aza-d-06-0-UI"
                      /clone_lib="NIH_BMAP_Ret3"
                      /dev_stage="6 weeks"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The
                      NIH_BMAP_Ret3 library is derived from mouse retina tissue.
                      For a detailed description of the library from which this
                      clone was derived, please visit our web site at
                      brainest.eng.uiowa.edu. The tissue for this library was
                      contributed by Dr. Xin-Yuan Fu, Yale University School of
                      Medicine
                      TAG_LIB=NIH_BMAP_Ret3
                      TAG_TISSUE=adult-retina
                      TAG_SEQ=GTCAGCGCGCAC"
BASE COUNT       90 a     86 c    126 g    100 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 12;  Length 402;
  Best Local Similarity   96.0%;  Pred. No. 19;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||||| |||||||||||||
Db    334 GCTTTCTTTGCTAAGAGCTCTTCCA 310


RESULT 2
BI737699
LOCUS       BI737699                 495 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603358627F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5365638 5',
            mRNA sequence.
ACCESSION   BI737699
VERSION     BI737699.1  GI:15714712
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11930  row: f  column: 07
            High quality sequence stop: 489.
FEATURES              Location/Qualifiers
     source           1. .495
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:5365638"
                      /clone_lib="NIH_MGC_94"
                      /tissue_type="retina"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                      Average insert size 3.3 kb. Library enriched for
                      full-length clones and constructed by Life Technologies.
                      Note: this is a NIH_MGC Library."
BASE COUNT      105 a    159 c    111 g    120 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 13;  Length 495;
  Best Local Similarity   96.0%;  Pred. No. 21;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||||| |||||||||||||
Db    270 GCTTTCTTTGCTAAGAGCTCTTCCA 294


RESULT 3
BI734160
LOCUS       BI734160                 628 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603351436F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5358776 5',
            mRNA sequence.
ACCESSION   BI734160
VERSION     BI734160.1  GI:15711173
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11912  row: h  column: 09
            High quality sequence start: 4
            High quality sequence stop: 627.
FEATURES              Location/Qualifiers
     source           1. .628
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:5358776"
                      /clone_lib="NIH_MGC_94"
                      /tissue_type="retina"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                      Average insert size 3.3 kb. Library enriched for
                      full-length clones and constructed by Life Technologies.
                      Note: this is a NIH_MGC Library."
BASE COUNT      159 a    180 c    152 g    137 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 13;  Length 628;
  Best Local Similarity   96.0%;  Pred. No. 23;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||||| |||||||||||||
Db     45 GCTTTCTTTGCTAAGAGCTCTTCCA 69
```

```
RESULT 4
BI736587
LOCUS       BI736587                636 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603361053F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5368044 5',
            mRNA sequence.
ACCESSION   BI736587
VERSION     BI736587.1  GI:15713600
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 636)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11936  row: j  column: 13
            High quality sequence stop: 632.
FEATURES             Location/Qualifiers
     source          1. .636
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5368044"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      145 a    208 c    147 g    136 t
ORIGIN

  Query Match            93.6%;  Score 23.4;  DB 13;  Length 636;
  Best Local Similarity  96.0%;  Pred. No. 24;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          |||||||||||| ||||||||||||
Db    179 GCTTTCTTTGCTAAGAGCTCTTCCA 203


RESULT 5
BG403992
LOCUS       BG403992                649 bp    mRNA    linear   EST 12-MAR-2001
DEFINITION  602419859F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4526966 5',
            mRNA sequence.
ACCESSION   BG403992
VERSION     BG403992.1  GI:13297440
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10435  row: e  column: 15
            High quality sequence stop: 644.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4526966"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      164 a    203 c    140 g    142 t
ORIGIN

  Query Match            93.6%;  Score 23.4;  DB 12;  Length 649;
  Best Local Similarity  96.0%;  Pred. No. 24;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          |||||||||||| ||||||||||||
Db    108 GCTTTCTTTGCTAAGAGCTCTTCCA 132


RESULT 6
BI730233
LOCUS       BI730233                651 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603350201F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5357472 5',
            mRNA sequence.
ACCESSION   BI730233
VERSION     BI730233.1  GI:15707246
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 651)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11909  row: b  column: 01
            High quality sequence stop: 539.
FEATURES             Location/Qualifiers
     source          1. .651
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5357472"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      154 a    198 c    149 g    150 t
```

```
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 13;  Length 651;
  Best Local Similarity   96.0%;  Pred. No. 24;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
           ||||||||||| ||||||||||||
Db      61 GCTTTCTTTGCTAAGAGCTCTTCCA 85


RESULT 7
BI730180
LOCUS       BI730180                 656 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603349724F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5357429 5',
            mRNA sequence.
ACCESSION   BI730180
VERSION     BI730180.1  GI:15707193
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 656)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11908  row: p  column: 06
            High quality sequence stop: 649.
FEATURES             Location/Qualifiers
     source          1. .656
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5357429"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      142 a    210 c    152 g    152 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 13;  Length 656;
  Best Local Similarity   96.0%;  Pred. No. 24;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
           ||||||||||| ||||||||||||
Db     402 GCTTTCTTTGCTAAGAGCTCTTCCA 426


RESULT 8
BI729678
LOCUS       BI729678                 663 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603349362F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5356959 5',
            mRNA sequence.
ACCESSION   BI729678
VERSION     BI729678.1  GI:15706691
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11907  row: l  column: 16
            High quality sequence stop: 659.
FEATURES             Location/Qualifiers
     source          1. .663
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5356959"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      156 a    209 c    151 g    147 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 13;  Length 663;
  Best Local Similarity   96.0%;  Pred. No. 24;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
           ||||||||||| ||||||||||||
Db     152 GCTTTCTTTGCTAAGAGCTCTTCCA 176


RESULT 9
BG298553
LOCUS       BG298553                 672 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602396937F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4511697 5',
            mRNA sequence.
ACCESSION   BG298553
VERSION     BG298553.1  GI:13063322
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 672)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10395  row: i  column: 10
            High quality sequence stop: 666.
FEATURES             Location/Qualifiers
     source          1. .672
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
```

```
                     /clone="IMAGE:4511697"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      150 a    206 c    156 g    160 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 12;  Length 672;
  Best Local Similarity   96.0%;  Pred. No. 24;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||||| ||||||||||||
Db    341 GCTTTCTTTGCTAAGAGCTCTTCCA 365


RESULT 10
BI736996
LOCUS       BI736996                676 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603360842F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5367875 5',
            mRNA sequence.
ACCESSION   BI736996
VERSION     BI736996.1  GI:15714009
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 676)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11936  row: c  column: 12
            High quality sequence stop: 662.
FEATURES             Location/Qualifiers
     source          1. .676
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5367875"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      144 a    214 c    160 g    158 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 13;  Length 676;
  Best Local Similarity   96.0%;  Pred. No. 24;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||||| ||||||||||||
Db    409 GCTTTCTTTGCTAAGAGCTCTTCCA 433


RESULT 11
BI735044
LOCUS       BI735044                681 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603356176F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5363395 5',
            mRNA sequence.
ACCESSION   BI735044
VERSION     BI735044.1  GI:15712057
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 681)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11924  row: h  column: 20
            High quality sequence stop: 680.
FEATURES             Location/Qualifiers
     source          1. .681
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5363395"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      152 a    220 c    156 g    152 t      1 others
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 13;  Length 681;
  Best Local Similarity   96.0%;  Pred. No. 24;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||||| ||||||||||||
Db    244 GCTTTCTTTGCTAAGAGCTCTTCCA 268


RESULT 12
BG295809
LOCUS       BG295809                682 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602393292F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4505134 5',
            mRNA sequence.
ACCESSION   BG295809
VERSION     BG295809.1  GI:13057815
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 682)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

```
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM10378  row: g  column: 23
           High quality sequence stop: 603.
FEATURES             Location/Qualifiers
     source          1. .682
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4505134"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      137 a    205 c    167 g    173 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 12;  Length 682;
  Best Local Similarity   96.0%;  Pred. No. 24;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||||| |||||||||||||
Db    512 GCTTTCTTTGCTAAGAGCTCTTCCA 536


RESULT 13
BI737008
LOCUS       BI737008                694 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603360855F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5368117 5',
           mRNA sequence.
ACCESSION   BI737008
VERSION     BI737008.1  GI:15714021
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 694)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM11936  row: m  column: 14
           High quality sequence stop: 694.
FEATURES             Location/Qualifiers
     source          1. .694
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5368117"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      157 a    216 c    156 g    165 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 13;  Length 694;
  Best Local Similarity   96.0%;  Pred. No. 24;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||||| |||||||||||||
Db    363 GCTTTCTTTGCTAAGAGCTCTTCCA 387


RESULT 14
BG298167
LOCUS       BG298167                696 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602396304F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4507824 5',
           mRNA sequence.
ACCESSION   BG298167
VERSION     BG298167.1  GI:13062367
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM10385  row: h  column: 01
           High quality sequence stop: 677.
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4507824"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      148 a    218 c    164 g    166 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 12;  Length 696;
  Best Local Similarity   96.0%;  Pred. No. 24;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
          ||||||||||||| |||||||||||||
Db    459 GCTTTCTTTGCTAAGAGCTCTTCCA 483


RESULT 15
BI735675
LOCUS       BI735675                697 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603357821F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5364964 5',
           mRNA sequence.
ACCESSION   BI735675
VERSION     BI735675.1  GI:15712688
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11928  row: j  column: 05
            High quality sequence stop: 697.
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5364964"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      136 a    209 c    168 g    184 t
ORIGIN

  Query Match             93.6%;  Score 23.4;  DB 13;  Length 697;
  Best Local Similarity   96.0%;  Pred. No. 24;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCTTTCTTTGCTGAGAGCTCTTCCA 25
           ||||||||||||| |||||||||||
Db     585 GCTTTCTTTGCTAAGAGCTCTTCCA 609


Search completed: March 17, 2003, 13:09:23
Job time : 886.737 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:43:42 ; Search time 202.495 Seconds
(without alignments)
3161.870 Million cell updates/sec

Title: US-09-836-439-6
Perfect score: 22
Sequence: 1 aagacttctgagtaacaatcaa 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		19.4	88.2	17938	3	AF198437	AF198437 Theileria
2		18.8	85.5	42208	9	AL589989	AL589989 Human DNA
3	c	18.8	85.5	116685	9	HS516C23	Z93021 Human DNA s
4	c	18.8	85.5	178168	2	AC012350	AC012350 Homo sapi
5		18.8	85.5	180970	9	AC009302	AC009302 Homo sapi
6		18.8	85.5	182152	2	AC067929	AC067929 Homo sapi
7		18.8	85.5	189281	2	AC012458	AC012458 Homo sapi
8		18.8	85.5	191434	9	AC087477	AC087477 Homo sapi
9	c	18.8	85.5	197437	2	AL772303	AL772303 Mus muscu
10		18.8	85.5	223761	2	AC127239	AC127239 Mus muscu
11		18.8	85.5	229247	10	AC098888	AC098888 Mus muscu
12		18.8	85.5	317515	2	AC099415	AC099415 Mus muscu
13	c	18.4	83.6	10243	1	AE011043	AE011043 Methanosa
14	c	18.4	83.6	10467	1	AE010860	AE010860 Methanosa
15		18.4	83.6	152556	9	CNS01DSY	AL122035 Human chr
16		18.4	83.6	153951	2	AC019027	AC019027 Homo sapi
17	c	18	81.8	148507	2	AC011263	AC011263 Homo sapi
18		18	81.8	158633	2	AC015472	AC015472 Homo sapi
19	c	18	81.8	178199	2	AC018772	AC018772 Homo sapi
20	c	18	81.8	224068	9	CNS01DUJ	AL133246 BAC seque
21	c	17.8	80.9	907	9	AF213436	AF213436 Homo sapi
22	c	17.8	80.9	927	6	I07695	I07695 Sequence 18
23	c	17.8	80.9	3159	9	HSDYSE51	X51934 Human dystr
24		17.8	80.9	57507	9	AL603652	AL603652 Human DNA
25		17.8	80.9	83007	9	AC097448	AC097448 Homo sapi
26		17.8	80.9	107406	2	AC013491	AC013491 Homo sapi
27	c	17.8	80.9	141984	8	F9L1	AC007591 Arabidops
28	c	17.8	80.9	148571	9	AC105316	AC105316 Homo sapi
29		17.8	80.9	161874	9	AC079864	AC079864 Homo sapi
30	c	17.8	80.9	163132	2	AC097180	AC097180 Rattus no
31	c	17.8	80.9	176287	10	AL513347	AL513347 Mouse DNA
32		17.8	80.9	179591	9	AC093902	AC093902 Homo sapi
33	c	17.8	80.9	182463	2	AC120773	AC120773 Rattus no
34		17.8	80.9	182933	2	AC097610	AC097610 Rattus no
35		17.8	80.9	184573	2	AC111144	AC111144 Mus muscu
36	c	17.8	80.9	191664	2	AC125751	AC125751 Rattus no
37	c	17.8	80.9	200539	2	AC121805	AC121805 Mus muscu
38		17.8	80.9	201296	10	AC115299	AC115299 Mus muscu
39	c	17.8	80.9	203753	9	AC104343	AC104343 Homo sapi
40		17.8	80.9	207945	2	AC117841	AC117841 Rattus no
41	c	17.8	80.9	207974	2	AL831771	AL831771 Mus muscu
42	c	17.8	80.9	208953	9	CNS01RH3	AL160314 Human chr
43	c	17.8	80.9	210779	9	AC007450	AC007450 Homo sapi
44		17.4	79.1	12024	1	AE000995	AE000995 Archaeogl
45	c	17.4	79.1	65499	2	AC100433	AC100433 Mus muscu

ALIGNMENTS

RESULT 1
AF198437
LOCUS AF198437 17938 bp DNA linear INV 05-DEC-2000
DEFINITION Theileria parva strain Muguga hypothetical telomeric SfiI fragment 20 protein 3, hypothetical telomeric SfiI fragment 20 protein 2, and hypothetical telomeric SfiI fragment 20 protein 1 genes, complete cds.
ACCESSION AF198437 L36964
VERSION AF198437.1 GI:11545211
KEYWORDS .
SOURCE Theileria parva.
ORGANISM Theileria parva
Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae; Theileria.
REFERENCE 1 (bases 17114 to 17938)

AUTHORS Sohanpal,B.K., Morzaria,S.P., Gobright,E.I. and Bishop,R.P.
TITLE Characterisation of the telomeres at opposite ends of a 3 Mb Theileria parva chromosome
JOURNAL Nucleic Acids Res. 23 (11), 1942-1947 (1995)
MEDLINE 95319937
PUBMED 7596822
REFERENCE 2 (bases 1 to 17938)
AUTHORS Bishop,R., Gobright,E., Nene,V., Morzaria,S., Musoke,A. and Sohanpal,B.
TITLE Polymorphic open reading frames encoding secretory proteins are located less than 3 kilobases from Theileria parva telomeres
JOURNAL Mol. Biochem. Parasitol. 110 (2), 359-371 (2000)
MEDLINE 20520968
PUBMED 11071289
REFERENCE 3 (bases 17114 to 17938)
AUTHORS Sohanpal,B.K., Morzaria,S.P., Gobright,E.I. and Bishop,R.P.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1995) Unit 5, International Livestock Research Institute (ILRI), Old Naivasha Road, Nairobi P.O. Box 30709, Kenya
REFERENCE 4 (bases 1 to 17938)
AUTHORS Bishop,R.P., Gobright,E.I. and Sohanpal,B.K.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-1999) Unit 5, International Livestock Research Institute (ILRI), Old Naivasha Road, Nairobi P.O. Box 30709, Kenya
COMMENT On Dec 5, 2000 this sequence version replaced gi:624245.
FEATURES Location/Qualifiers

source 1. .17938
/organism="Theileria parva"
/strain="Muguga"
/db_xref="taxon:5875"
/chromosome="1"
/map="SfiI fragment 20"

CDS 251. .10840
/note="ORF3"
/codon_start=1
/product="hypothetical telomeric SfiI fragment 20 protein 3"
/protein_id="AAG37800.1"
/db_xref="GI:11545212"
/translation="MRKCMIFIKLLFYIALLHKWNGVESQPTSSGRSNSSGSQGSRAQ
STTPKSQNQTQSSQSQSESTASATSRRRTGTTGAQGSQPTKPPSQPASSNGSGSHASP
RASTGQSSTPKSQPTTPTSQSQPKSTASRTSSTGVGQARGTPSSATSQPSRGVGVGRG
SQPKATAATIGTSTPTTPPTPTTSAAQTKATPTTGARPATQPKVTPASASSTQPKSQA
NVTPVSGSAAPSVSTSATPTKPVSQSTPTTSVTKVTLNIDTTKESTSELDYSKIGDIH
NYTPKQNHLFSKVTHGSVVVWESKTEVHGTRVTYVCIIKYLVVLLSNDMFLLFHQSSG
TWTDVTKERYDVDKLKFFGENDAALNKSSYSTTLISYSYSLEFNAGVKCQKVKYGNED
VWKHSDDTDFGSIKSLFIDLISNQLSVKNESGQLKKLEYKPKESNGTTSETADATKEP
TTPTTVTTKSDPTTTATATETSSSTLTTESDTATTQTTPVTTPTSSGSSVTQSKRLES
SSTSGNIHTFTSGVSHSAVDASQGASVLDSGSSSSTLVLTEYKRLPSIRLTTSTDTDT
SSKSQTATPVTTTTTNVALDIDKKQSTDQFEYSDINNYRIYVAKTNFVFNKIVQGTTV
IWESKDYCATSVTTKTIDNKVFLAIVLSNGMFTLVEKSSGKWTDITLIGTWYLHLKFY
GDRDTEIKSSKYSVTMSGLSYNLDFNTNVDCKKIKYNNVTLWSHTDDSEFGTIKSLSL
DLAKNVFFVKNSADIVKKLELKYISLDIDKKQSTDESDYSKSDCHTYMTKSNFVFNRI
VQGTDFVWKSQVLFGTKVTAAGDFLTILLNNGTFKLFQKSVGEWIDITPNKHDIGKLK
FYGEGDRLLKSYDYSVTLSCLAFSYEFNVGSECHKVKYGDDVLWNHSDDSQFGTIKTF
SLDLPKNQFTIKNNSGTVKNVEFKYVILDLNTMESTDHFDRTGDDGIFKFVVKSGYVF
SKVVQGTSVVWEPRDSAYCTEVTYIVAGDFKYLSAVLDNGSCLLLYKTSKDKPFKDIT
STSYNVTRLKFLTDDNSELSTSDYKVTVAEHCYNLEFNSDVKCSKITHNGADLWKHSD
DTQFGTIKSLHLNLINNQLGLKNQSNTLKELQAPLPPTPDVTDTTGTKITIDLNETKE
VEGMFTYSHDGKCHTYATAGFGNTFGKVMKGTRVVWEPVGNVQAKKIIYVDDHVKNVV
VVLTDGKFLLLKQSDNNWDDVTKYTLDVSKLVFLGDNHGATSLGVSASVTVVDTAMTL
TPDDQSKLISGSASSSPVTSPVSQPKSATVTTTPATTPQTQGPAPAKVPEAKSTTTTA
PQVSGPAVTPSTATPSPAKVTSPPAATLSGTEVTPPKVAVSGTPVTLDIDNTQSASEY
DYKDEGGVVTYTPKENHVFSKVTEGPIVLWKSDDVSGTMVRICSKLLVILFNNNEFKL
FEQSEGNWIDITDKRHNLTKLKFFNDSNAQIKPTEYTVTIVELSYRIQFSQGTICRKI
TLGNDIIYNYQDDTDFGSIGSLDLDLVSNGFFVKNESGDLKSVTKPVILDLETTKSTP
EFDISYIEGSDHIYIPKENRVFNKVVRGPTLIWESKDNVCATIVTISSKFLSILFDNN
TFKLFQESEKQWTDITSDRHNVASLKFYGDNNVEIKLSDYIVSMSGLSYNFEFKTGVK
CKKITLDNVTLWSHTDDTQFGTIKSFSLGLVLNGFFIKNESGTVKPAEPTKVTLDIEK
TQSTVQLEYKVQNGVVTYTGKDCHVFSKIMQGSTVVWESEENLFATKVIKDGNKNSLS
INLSNGTTKEFQQSDSKWTESQTSSPVTTSSSSAQTATTETATTGTTTTTKTVLDIDI
KKSTDQFQYSNSGGFTKFVARPGYGFTKVTKGNSVVFQSTRILFGTIVTYMDSCNYVS
AQFINGMFKDFQLSGTKWTESKPIQSNNTPFTSFDYKVTVTNYIYEFQFNEGAECKTV
KYGNVQVYRHTDHPFQRSFGSIKSFSVDLVKNRFFIKNQSCSMRELPYEPPLDTKSID
TTQRLLLTRGQSTTDVIEAVDVRPITLDLTTATNNKFSRTDKDGYVTYTANPGFVFGK
VTHGANIVWESDSNIFGNLVRIKSESNNMFLVILLDNSTFKLFQFSDGTVFKDITAEK
RDVTKLKFLGEGGEINPSDFKVSLADFQFSYEFHDRVKCLRITLGNECFYEHKGGFGA
IKSLSLDLVLNKFSTVEVKSVSLDIDKSQSTNEFDHTEKNGVSTFSPKSGHVFTKVTH
SNTAIWESDGNVCATIVNTKTSDNKVFLSIFLSNNMFKLFLKSSDGAFKDITNQRHDV
KKLKFLGQDDSPLKSTDYSVSMSGHTFLCEFNRGFNCLGVKLGNEDIWKPDGDPKFST
IKSLSLDLIDNKFSLKNESGDIKPVVIPVILDISKTQSTSEFEYLKEYEFHKYTVKSC
YLFTKVTEGAKVVWESNDTVGGTLVTTTGEFLAILLNNNLFKLFLKSSEWEDITKSRI
DISGLKFLTDDSEIGTSDYKVTISCHSYILEFNPGVECHKITHRGIDLWKSDDDQKFK
SVKSLSLCLLSNQLSVTNESGTVKDLKLTSDTTETESTAPVTKYITLDVDKKQSTSEF
ECKEKNGVTSYTAKSDIVFTKITHKTTVVWESGGNVCAILVSVKAEFLAILLSNNQFK
LFQKSSDGTWKDITSQRHNISNLKFLGQGGSVLKSPDYSVSMSGLILVYEFNTKCTKI
MLENENIWKPDDDPKFPKIKSLSLDIASNKFSLKSDTAVKELELSLATSTESATKSAT
ATSVTDPTTQTTPESVTGFPTKSSPESTTTTQPNVSTTTQSVTPTPTGTTVTSVTLDI
NTTNSTQQFECKEKNGVISYVAKSGNVFTKVTHGQTSVWESSGNVSGAMVTTKKDEFI
AILLDNKKFLIFQNSDNKWTDVTCDRCDLDELSFFSDDTVLTSSQYKIVLSDLSYNLQ
FNDGVKCSKITHKGVRVWPVPDGSKTTTDTDDTTKTIDDTTSSKSASDDTGFGFIKSL
HLDLVANKFSVKNESDQVKEFEYKPVKSKIITSSSGGTLSVDAVPGLYKLNLGLTQSQ
YDIVKVGKSFEYHFKPNVKCSMILFNDDTTSYYHPTVWNSPNQDDVFGTKVVVSDDVK
YLLLLLNNNMYKMFTHFYNRWVEISRDRRDVSKLKFYGIQEVSDSSSSLTVPESDRQL
TPSDYKFSISDFVFSFEFNEGILCRKINLGDMIVYNYDILKTPIKSFNINLRTNQFSV
NNRFGDANVELVQRPDTLITIDLNVCETTFEYQYLKTDDDFHTYVPLFGYAFNKIVRG
SNDIWESTTDCSLKVVYMDSAKLLAILMDNNKFFLFQESSGTFRDITSMRHDVSNLKF
LDTRDSELTSSDLKITIVNNSYRYEFKDGVRCQTVKYSDDVLWRHSDHTKFGTIVSFH
LGLVSNKFFVKNQSDQLKMIEYRPGTLPIVTDVNRNTSSELDIDKKGSTDRYDYENNG
DDHIFTPKSGCEFSKVVRGEKVIWEASENLFAKDVIFDEANNYVTIELTNGSFRTITL
SDTTT"

CDS 11463. .12878
/note="ORF2"
/codon_start=1
/evidence=not_experimental
/product="hypothetical telomeric SfiI fragment 20 protein 2"
/protein_id="AAG37801.1"
/db_xref="GI:11545213"
/translation="MTAKWLVINLLLYIRLLCQLNSVVAHPTSGTRSPTNGFQTSRGQ
YSTPNGQTTFPEINVTRVSTPASNSTKQVTLDVDKTGSTNDYEYSKENKDNKYHTYTV
KQGNVITKVTKGNAVMWDSKNDDYCLKAIYMDHHNSPKYLAVLLQSGKFVLIYRAAYD
VIERSDTNMNITHQRYDVTKLKFLGDSDAVIDASSYEPTIADRYYSFEFKNGAKCKKI
KFGDDVLWKHDSEFGQIVSFHLDLISNKFFLKNHNGDVKSIEYTTGSLKVTTVTTNCL
LLPTPESTTVSPELTTLAPESTTVIPDQTETTATFAESTSSYNPSTLDLNTRNSTRDF
DYSKTDNVHVYTAKSNHSFSKITRGSNTVWEAPDNVSAKSVIFVDESIKYVTIDLGGG
FYRTVQVDPKSQTNTKTNNTRTKNIPLILLVPLILLPPTLLTLLILLLKHLLLLKKQ
KQLLKLEQPLPLNQQMLLPHM"

CDS 13341. .15017
/note="ORF1"
/codon_start=1
/evidence=not_experimental
/product="hypothetical telomeric SfiI fragment 20 protein 1"
/protein_id="AAG37802.1"
/db_xref="GI:11545214"
/translation="MIGHWLFINLLLYITLLCQMNSVVTYSISSTSAKQITLDINKTT
STNDSDYSKNGKYRTFTAKPGKSFNKVVKKKLYIWECKNNDQAVKVVLMGSGNNKKHL
LILLKSSNFVILYKTGRNKPWTNITSKRHDPTNLKFFDIDKELTSSDYEVTIYEFSYR
FTFKAGVNCKRINYGDVVIWTHTDDKFEYITNFDLDLVKNEFFVEDDKRKSRKVNYEG
PIISSVTPRTASASTPSTPKYQYASATKVTLDISKTQFTCFEYSLKSGVIRTFTPRSN
CFIIKVTHCTTVIWESKTDVFGIRVTYVDVVKYLFLLLDNDKFLLFHQPPDRNKPWID
ITYTRVNLSKLKFFAENDMELKSTDYSVNLVSYSYSLEFNAGVKCQKVKYGNEDVWTN
SDDTDFGYIKSLLIDLVANKFRVKNDSGQLKKLDFKPSTPRLIPIPEQQYQQAPPQSR
PTPTYPRPPIPARPPYPSHQLPVQPQPIHPGPQLSVRQPVRPPAYSEIQYDTPETPSY
PPPPYSESQHDTPETPSYPPPPYSESQHDTPEPPSYPPPPYSESQHEAPQ"

repeat_region 17698. .17938
/note="telomeric repeat"
/function="stabilizes end of chromosome"
/rpt_family="telomeric"
/rpt_type=tandem
/rpt_unit=17698. .17715

BASE COUNT 5477 a 3549 c 3098 g 5814 t
ORIGIN

Query Match 88.2%; Score 19.4; DB 3; Length 17938;

```
  Best Local Similarity   95.2%;  Pred. No. 56;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 AAGACTTCTGAGTAACAATCA 21
          || ||||||||||||||||||
Db   8185 AATACTTCTGAGTAACAATCA 8205


RESULT 2
AL589989
LOCUS       AL589989               42208 bp    DNA     linear   PRI 12-MAY-2001
DEFINITION  Human DNA sequence from clone RP1-116B8 on chromosome 6, complete
            sequence.
ACCESSION   AL589989
VERSION     AL589989.6  GI:14041769
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 42208)
  AUTHORS   Tromans,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 15, 2001 this sequence version replaced gi:13751581.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP1-116B8 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP1-116B8 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-13J16 is at 42109 in this sequence.
            The true right end of clone RP3-483L3 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .42208
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP1-116B8"
                     /clone_lib="RPCI-1"
     repeat_region   546. .923
                     /note="MLT1B repeat: matches 1. .383 of consensus"
     repeat_region   1645. .1739
                     /note="MIR repeat: matches 156. .262 of consensus"
     repeat_region   2339. .2368
                     /note="15 copies 2 mer tt 86% conserved"
     repeat_region   3125. .3190
                     /note="33 copies 2 mer ca 66% conserved"
     repeat_region   3464. .3770
                     /note="AluY repeat: matches 1. .309 of consensus"
     repeat_region   3996. .4294
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   4295. .4364
                     /note="MER44 repeat: matches 655. .728 of consensus"
     repeat_region   4374. .6566
                     /note="TIGGER1 repeat: matches 2. .2227 of consensus"
     repeat_region   6823. .7007
                     /note="AluY repeat: matches 116. .307 of consensus"
     repeat_region   7013. .7190
                     /note="TIGGER1 repeat: matches 2241. .2418 of consensus"
     repeat_region   7450. .7711
                     /note="MER44A repeat: matches 2. .238 of consensus"
     repeat_region   7988. .8091
                     /note="MIR repeat: matches 41. .146 of consensus"
     repeat_region   8780. .8964
                     /note="L1MA9 repeat: matches 6087. .6270 of consensus"
     repeat_region   9608. .9657
                     /note="AluJ/FLAM repeat: matches 10. .59 of consensus"
     repeat_region   10806. .10945
                     /note="MIR repeat: matches 13. .148 of consensus"
     repeat_region   11078. .11372
                     /note="AluJo repeat: matches 4. .298 of consensus"
     repeat_region   12599. .13021
                     /note="L1MB3 repeat: matches 5737. .6184 of consensus"
     repeat_region   13039. .13314
                     /note="AluJo repeat: matches 7. .298 of consensus"
     repeat_region   14353. .14411
                     /note="L2 repeat: matches 2405. .2459 of consensus"
     repeat_region   14412. .14724
                     /note="AluJo repeat: matches 1. .301 of consensus"
     repeat_region   14725. .14934
                     /note="L2 repeat: matches 2459. .2691 of consensus"
     repeat_region   16429. .16820
                     /note="MER63A repeat: matches 44. .208 of consensus"
     repeat_region   17156. .17192
                     /note="MER63A repeat: matches 1. .37 of consensus"
     repeat_region   17998. .18335
                     /note="L2 repeat: matches 2158. .2506 of consensus"
     repeat_region   18384. .18689
                     /note="MER2 repeat: matches 22. .345 of consensus"
     repeat_region   18690. .19270
                     /note="LTR18B repeat: matches 1. .603 of consensus"
     repeat_region   19467. .20027
                     /note="L1MA4A repeat: matches 5717. .6288 of consensus"
     repeat_region   20646. .20788
                     /note="L2 repeat: matches 2158. .2302 of consensus"
     repeat_region   21102. .21859
                     /note="L1MB3 repeat: matches 5434. .6176 of consensus"
     repeat_region   22150. .22356
                     /note="L2 repeat: matches 1246. .1462 of consensus"
     repeat_region   22396. .22600
                     /note="L2 repeat: matches 112. .316 of consensus"
     repeat_region   22706. .22754
                     /note="MER3 repeat: matches 152. .201 of consensus"
     repeat_region   22755. .22883
                     /note="FLAM_A repeat: matches 1. .130 of consensus"
     repeat_region   22884. .23038
                     /note="MER3 repeat: matches 3. .152 of consensus"
     repeat_region   23345. .23433
                     /note="L1PA13 repeat: matches 6063. .6156 of consensus"
     repeat_region   23561. .23632
                     /note="36 copies 2 mer at 65% conserved"
     repeat_region   23694. .23989
                     /note="MLT1A1 repeat: matches 44. .315 of consensus"
     repeat_region   24062. .24213
                     /note="MER5B repeat: matches 3. .173 of consensus"
     repeat_region   25110. .25213
                     /note="MIR repeat: matches 51. .157 of consensus"
     repeat_region   26282. .26447
                     /note="L2 repeat: matches 2482. .2626 of consensus"
     repeat_region   26454. .26782
```

```
                     /note="MLT1A1 repeat: matches 7. .365 of consensus"
     repeat_region   26935. .27440
                     /note="MER67A repeat: matches 1. .542 of consensus"
     repeat_region   28802. .28943
                     /note="2 copies 71 mer 88% conserved"
     repeat_region   29336. .29743
                     /note="MLT1J repeat: matches 1. .516 of consensus"
     repeat_region   30888. .31049
                     /note="MIR repeat: matches 91. .255 of consensus"
     repeat_region   31296. .31356
                     /note="MIR repeat: matches 126. .190 of consensus"
     repeat_region   32851. .32963
                     /note="MIR repeat: matches 80. .190 of consensus"
     repeat_region   33104. .33456
                     /note="MSTD repeat: matches 1. .394 of consensus"
     repeat_region   34340. .34703
                     /note="L1MC/D repeat: matches 5395. .5756 of consensus"
     repeat_region   34744. .34975
                     /note="MIR repeat: matches 23. .261 of consensus"
     repeat_region   35717. .35886
                     /note="MIR repeat: matches 34. .225 of consensus"
     repeat_region   36532. .36728
                     /note="MIR repeat: matches 28. .232 of consensus"
     repeat_region   37830. .37923
                     /note="47 copies 2 mer ca 72% conserved"
     repeat_region   37925. .37982
                     /note="MER4-internal repeat: matches 1111. .1162 of
                     consensus"
     repeat_region   37990. .38133
                     /note="72 copies 2 mer ca 74% conserved"
     repeat_region   39553. .39590
                     /note="19 copies 2 mer ac 100% conserved"
     repeat_region   39893. .39971
                     /note="MIR repeat: matches 66. .147 of consensus"
     repeat_region   39973. .41078
                     /note="L1M2 repeat: matches 4729. .5799 of consensus"
     repeat_region   41106. .41153
                     /note="L1MB1 repeat: matches 5782. .5829 of consensus"
     repeat_region   41840. .41963
                     /note="LTR16C repeat: matches 259. .387 of consensus"
BASE COUNT    12256 a    9645 c    9179 g   11128 t
ORIGIN

  Query Match             85.5%;  Score 18.8;  DB 9;  Length 42208;
  Best Local Similarity   90.9%;  Pred. No. 1e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGACTTCTGAGTAACAATCAA 22
           || |||||||||||||| |||
Db   33877 AACACTTCTGAGTAACAAACAA 33898


RESULT 3
HS516C23/c
LOCUS       HS516C23              116685 bp    DNA     linear   PRI 23-NOV-1999
DEFINITION  Human DNA sequence from clone 516C23 on chromosome 6q12 Contains CA
            repeat (D6S402) and GSSs, complete sequence.
ACCESSION   Z93021
VERSION     Z93021.2  GI:4539080
KEYWORDS    HTG; repeat polymorphism.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 116685)
  AUTHORS   Milne,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 29, 1999 this sequence version replaced gi:3821035.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence is the entire insert of clone 516C23. This sequence
            has been finished according to sequence map criteria as follows. An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 6, constructed by the Sanger Centre Chromosome 6
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            516C23 is from the library RPCI3 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .116685
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="q12"
                     /clone="RP3-516C23"
                     /clone_lib="RPCI-3"
     repeat_region   1. .2983
                     /note="L1PA12 repeat: matches 751. .3796 of consensus"
     repeat_region   3221. .3704
                     /note="L1PA12 repeat: matches 252. .740 of consensus"
     repeat_region   3656. .4127
                     /note="L1PA12 repeat: matches 131. .587 of consensus"
     repeat_region   4135. .4260
                     /note="L1PA12 repeat: matches 534. .659 of consensus"
     repeat_region   4359. .4701
                     /note="THE1B repeat: matches 15. .364 of consensus"
     repeat_region   4779. .5222
                     /note="L1PA12 repeat: matches -1411. .-954 of consensus"
     repeat_region   5223. .5418
                     /note="MLT2E repeat: matches 8. .201 of consensus"
     repeat_region   5676. .5798
                     /note="LTR16C repeat: matches 260. .387 of consensus"
     repeat_region   7126. .7224
                     /note="MIR repeat: matches 166. .262 of consensus"
     repeat_region   9021. .9292
                     /note="L2 repeat: matches 2381. .2655 of consensus"
     repeat_region   9295. .9727
                     /note="L1MB8 repeat: matches 5015. .5445 of consensus"
     repeat_region   9728. .10090
                     /note="MLT1A1 repeat: matches 1. .365 of consensus"
     repeat_region   10094. .10793
                     /note="L1MB8 repeat: matches 5467. .6169 of consensus"
     repeat_region   10894. .12058
                     /note="L1PA2 repeat: matches 4981. .6144 of consensus"
     repeat_region   12061. .12110
                     /note="25 copies 2 mer aa 78% conserved"
     repeat_region   12705. .12748
                     /note="L2 repeat: matches 2655. .2698 of consensus"
     repeat_region   14121. .14150
                     /note="15 copies 2 mer gt 100% conserved"
     repeat_region   14401. .14782
                     /note="MLT1B repeat: matches 1. .390 of consensus"
     repeat_region   16400. .16486
                     /note="L2 repeat: matches 2657. .2746 of consensus"
     repeat_region   16859. .16985
                     /note="L2 repeat: matches 2570. .2698 of consensus"
     repeat_region   18906. .19147
                     /note="L2 repeat: matches 1706. .1969 of consensus"
     repeat_region   19219. .19457
                     /note="MIR repeat: matches 9. .248 of consensus"
     repeat_region   20472. .20676
                     /note="L2 repeat: matches 627. .842 of consensus"
     repeat_region   20803. .21698
```

```
                     /note="L1MC1 repeat: matches 5398. .6314 of consensus"
     repeat_region   23343. .23754
                     /note="L2 repeat: matches 2029. .2520 of consensus"
     repeat_region   24063. .24264
                     /note="MLT1B repeat: matches 189. .390 of consensus"
     repeat_region   24302. .24345
                     /note="22 copies 2 mer tt 75% conserved"
     repeat_region   24361. .30507
                     /note="L1PA2 repeat: matches 4. .6144 of consensus"
     repeat_region   30510. .30679
                     /note="MLT1B repeat: matches 1. .170 of consensus"
     repeat_region   31272. .31782
                     /note="L2 repeat: matches 960. .2750 of consensus"
     repeat_region   31811. .32018
                     /note="MIR repeat: matches 39. .262 of consensus"
     repeat_region   32300. .33151
                     /note="L1PB1 repeat: matches 5323. .6155 of consensus"
     repeat_region   33152. .33236
                     /note="AluSp/q repeat: matches 192. .276 of consensus"
     repeat_region   33484. .33868
                     /note="L2 repeat: matches 1862. .2267 of consensus"
     repeat_region   33887. .34010
                     /note="LTR33 repeat: matches 1. .143 of consensus"
     repeat_region   34187. .34218
                     /note="16 copies 2 mer aa 84% conserved"
     repeat_region   34220. .34364
                     /note="LTR33 repeat: matches 382. .517 of consensus"
     repeat_region   34403. .34964
                     /note="L1MEc repeat: matches 272. .906 of consensus"
     repeat_region   35050. .35158
                     /note="L1MA4 repeat: matches 6176. .6290 of consensus"
     repeat_region   35141. .35266
                     /note="L1M4 repeat: matches 2121. .2241 of consensus"
     repeat_region   35267. .35587
                     /note="AluY repeat: matches 1. .309 of consensus"
     repeat_region   35588. .36308
                     /note="L1M4 repeat: matches 2241. .2996 of consensus"
     repeat_region   36248. .36925
                     /note="L1ME2 repeat: matches 5476. .6154 of consensus"
     repeat_region   36950. .37121
                     /note="L1MEc repeat: matches 1782. .1975 of consensus"
     repeat_region   37125. .37741
                     /note="L1M4 repeat: matches 2359. .3033 of consensus"
     repeat_region   37749. .37877
                     /note="AluSq/x repeat: matches 1. .129 of consensus"
     repeat_region   37878. .37923
                     /note="23 copies 2 mer ca 100% conserved"
     repeat_region   37928. .38256
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   38262. .38434
                     /note="L1M4 repeat: matches 3047. .3235 of consensus"
     repeat_region   38463. .38837
                     /note="L1M4 repeat: matches 3790. .4186 of consensus"
     repeat_region   38847. .39138
                     /note="AluJo repeat: matches 12. .293 of consensus"
     repeat_region   39344. .39389
                     /note="23 copies 2 mer aa 76% conserved"
     repeat_region   40977. .41071
                     /note="MIR repeat: matches 12. .120 of consensus"
     repeat_region   41529. .41819
                     /note="L2 repeat: matches 1780. .2071 of consensus"
     misc_feature    41800. .41805
                     /note="IS2 transposable element excised from this
                     position"
     repeat_region   41820. .42230
                     /note="MLT1J repeat: matches 56. .505 of consensus"
     repeat_region   42267. .42627
                     /note="L2 repeat: matches 2301. .2677 of consensus"
     repeat_region   42644. .42731
                     /note="44 copies 2 mer aa 69% conserved"
     repeat_region   43499. .43640
                     /note="71 copies 2 mer tt 63% conserved"
     repeat_region   44092. .44158
                     /note="MLT1D repeat: matches 1. .66 of consensus"
     repeat_region   44166. .44649
                     /note="MLT1D repeat: matches 5. .505 of consensus"
     repeat_region   44686. .45053
                     /note="MLT1D repeat: matches 113. .503 of consensus"
     repeat_region   45793. .45960
                     /note="MIR repeat: matches 7. .171 of consensus"
     repeat_region   46079. .46271
                     /note="MLT1F repeat: matches 347. .541 of consensus"
     repeat_region   46316. .46363
                     /note="MLT2CB repeat: matches 236. .283 of consensus"
     repeat_region   46602. .46641
                     /note="20 copies 2 mer tg 95% conserved"
     misc_feature    complement(47028. .47443)
                     /note="match: GSS B53031 clone 2008G6"
     repeat_region   47509. .48290
                     /note="L1MB5 repeat: matches 5385. .6166 of consensus"
     repeat_region   48283. .49518
                     /note="L1MB5 repeat: matches 3964. .5271 of consensus"
     repeat_region   49519. .49586
                     /note="34 copies 2 mer ta 72% conserved"
     repeat_region   49980. .50288
                     /note="AluSq repeat: matches 3. .311 of consensus"
     repeat_region   50661. .51118
                     /note="L1MC4 repeat: matches 6102. .7977 of consensus"
     misc_feature    complement(51869. .52243)
                     /note="match: GSS AQ109033 clone 2377A15"
     repeat_region   52215. .52242
                     /note="14 copies 2 mer ac 100% conserved"
     misc_feature    <52243. .>52703
                     /note="match: GSS AQ234004"
     repeat_region   54296. .54852
                     /note="L2 repeat: matches 2180. .2743 of consensus"
     repeat_region   54942. .55097
                     /note="MIR repeat: matches 49. .210 of consensus"

  Query Match             85.5%;  Score 18.8;  DB 9;  Length 116685;
  Best Local Similarity   90.9%;  Pred. No. 90;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGACTTCTGAGTAACAATCAA 22
           ||  ||||||||||||||||||
Db   59346 AACTCTTCTGAGTAACAATCAA 59325


RESULT 4
AC012350/c
LOCUS       AC012350              178168 bp    DNA     linear   HTG 01-APR-2000
DEFINITION  Homo sapiens clone RP11-16N9, WORKING DRAFT SEQUENCE, 16 unordered
            pieces.
ACCESSION   AC012350
VERSION     AC012350.3  GI:7381803
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178168)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-16N9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178168)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
```

```
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 1, 2000 this sequence version replaced gi:6479001.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3593
                Center clone name: 16_N_9
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 172070 bases at least Q40
                Consensus quality: 174479 bases at least Q30
                Consensus quality: 175597 bases at least Q20
                Insert size: 170000; agarose-fp
                Insert size: 176668; sum-of-contigs
                Quality coverage: 6.4 in Q20 bases; agarose-fp
                Quality coverage: 6.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1694: contig of 1694 bp in length
            *     1695 1794: gap of      100 bp
            *     1795     4437: contig of 2643 bp in length
            *     4438 4537: gap of      100 bp
            *     4538     6644: contig of 2107 bp in length
            *     6645 6744: gap of      100 bp
            *     6745    10092: contig of 3348 bp in length
            *    10093 10192: gap of      100 bp
            *    10193    11801: contig of 1609 bp in length
            *    11802 11901: gap of      100 bp
            *    11902    15584: contig of 3683 bp in length
            *    15585 15684: gap of      100 bp
            *    15685    19675: contig of 3991 bp in length
            *    19676 19775: gap of      100 bp
            *    19776    27913: contig of 8138 bp in length
            *    27914 28013: gap of      100 bp
            *    28014    34156: contig of 6143 bp in length
            *    34157 34256: gap of      100 bp
            *    34257    43325: contig of 9069 bp in length
            *    43326 43425: gap of      100 bp
            *    43426    55265: contig of 11840 bp in length
            *    55266 55365: gap of      100 bp
            *    55366    68306: contig of 12941 bp in length
            *    68307 68406: gap of      100 bp
            *    68407    82910: contig of 14504 bp in length
            *    82911 83010: gap of      100 bp
            *    83011   102181: contig of 19171 bp in length
            *   102182 102281: gap of      100 bp
            *   102282   128058: contig of 25777 bp in length
            *   128059 128158: gap of      100 bp
            *   128159   178168: contig of 50010 bp in length.
FEATURES            Location/Qualifiers
     source         1. .178168
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="RP11-16N9"
                    /clone_lib="RPCI-11 Human Male BAC"
     misc_feature   1. .1694
                    /note="assembly_fragment"
     misc_feature   1795. .4437
                    /note="assembly_fragment"
     misc_feature   4538. .6644
                    /note="assembly_fragment"
     misc_feature   6745. .10092
                    /note="assembly_fragment"
     misc_feature   10193. .11801
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:right"
     misc_feature   11902. .15584
                    /note="assembly_fragment"
     misc_feature   15685. .19675
                    /note="assembly_fragment"
     misc_feature   19776. .27913
                    /note="assembly_fragment"
     misc_feature   28014. .34156
                    /note="assembly_fragment"
     misc_feature   34257. .43325
                    /note="assembly_fragment"
     misc_feature   43426. .55265
                    /note="assembly_fragment"
     misc_feature   55366. .68306
                    /note="assembly_fragment"
     misc_feature   68407. .82910
                    /note="assembly_fragment"
     misc_feature   83011. .102181
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:right"
     misc_feature   102282. .128058
                    /note="assembly_fragment"
     misc_feature   128159. .178168
                    /note="assembly_fragment"
BASE COUNT    50059 a  39406 c  38562 g  48641 t   1500 others
ORIGIN

  Query Match             85.5%;  Score 18.8;  DB 2;  Length 178168;
  Best Local Similarity   90.9%;  Pred. No. 86;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGACTTCTGAGTAACAATCAA 22
           || |||||||||||||||| |||
Db   16379 AACACTTCTGAGTAACAAACAA 16358


RESULT 5
AC009302
LOCUS       AC009302              180970 bp    DNA     linear   PRI 30-SEP-2000
DEFINITION  Homo sapiens BAC clone RP11-71J24 from 2, complete sequence.
ACCESSION   AC009302
VERSION     AC009302.2  GI:9931944
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180970)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 180970)
  AUTHORS   Tomlinson,C., Wohldmann,P., Maupin,R. and Reitz,L.
  TITLE     The sequence of Homo sapiens BAC clone RP11-71J24
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 180970)
  AUTHORS   Waterston,R.H.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 180970)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 180970)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 180970)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Aug 28, 2000 this sequence version replaced gi:5732171.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0071J24
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-350I24; the clone sequenced
            to the right is RP11-296A19.  Actual start of this clone is at base
            position 1 of RP11-71J24; actual end is at base position 180970 of
            RP11-71J24.
FEATURES            Location/Qualifiers
     source         1. .180970
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="2"
                    /map="2"
                    /clone="RP11-71J24"
                    /clone_lib="RPCI-11"
     repeat_region  4266. .4361
                    /rpt_family="MIR"
     repeat_region  7373. .7538
                    /rpt_family="MIR"
     repeat_region  7878. .8175
                    /rpt_family="Alu"
     repeat_region  8215. .8281
                    /rpt_family="MaLR"
     repeat_region  8699. .8778
                    /rpt_family="MIR"
     repeat_region  9014. .9113
                    /rpt_family="L1"
     repeat_region  9120. .9530
                    /rpt_family="L1"
     repeat_region  9639. .9748
                    /rpt_family="L2"
     repeat_region  11531. .11729
                    /rpt_family="MIR"
     repeat_region  11735. .11888
                    /rpt_family="L1"
     repeat_region  12416. .12459
                    /rpt_family="L2"
     repeat_region  12608. .12966
                    /rpt_family="MaLR"
     repeat_region  14055. .14365
                    /rpt_family="Alu"
     repeat_region  14659. .14822
                    /rpt_family="MIR"
     repeat_region  15167. .15466
                    /rpt_family="Alu"
     repeat_region  15865. .15942
                    /rpt_family="L2"
     repeat_region  16226. .16552
                    /rpt_family="Alu"
     repeat_region  17132. .17438
                    /rpt_family="Alu"
     repeat_region  19203. .19508
                    /rpt_family="Alu"
     repeat_region  19511. .19818
                    /rpt_family="Alu"
     repeat_region  20445. .20594
                    /rpt_family="Alu"
     repeat_region  20732. .21129
                    /rpt_family="ERVL"
     repeat_region  22146. .22254
                    /rpt_family="MER1_type"
     repeat_region  23476. .23592
                    /rpt_family="MIR"
     repeat_region  24585. .24914
                    /rpt_family="L1"
     repeat_region  25584. .25744
                    /rpt_family="MIR"
     repeat_region  25798. .26320
                    /rpt_family="MER1_type"
     repeat_region  29905. .30145
                    /rpt_family="MIR"
     repeat_region  30933. .31063
                    /rpt_family="ERVL"
     repeat_region  31064. .31118
                    /rpt_family="ERVL"
     repeat_region  31135. .31192
                    /rpt_family="ERVL"
     repeat_region  31213. .31623
                    /rpt_family="ERVL"
     repeat_region  31690. .32052
                    /rpt_family="MaLR"
     repeat_region  32252. .32408
                    /rpt_family="MIR"
     repeat_region  32638. .32993
                    /rpt_family="MaLR"
     repeat_region  33136. .33707
                    /rpt_family="MER2_type"
     repeat_region  35611. .35898
```

```
                     /rpt_family="Alu"
     repeat_region   37177. .37310
                     /rpt_family="MIR"
     repeat_region   37425. .37718
                     /rpt_family="Alu"
     repeat_region   38762. .39081
                     /rpt_family="Alu"
     repeat_region   39123. .39159
                     /rpt_family="L2"
     repeat_region   39234. .39560
                     /rpt_family="MER1_type"
     repeat_region   39561. .39809
                     /rpt_family="L2"
     repeat_region   39810. .40188
                     /rpt_family="MaLR"
     repeat_region   40189. .40265
                     /rpt_family="L2"
     repeat_region   40266. .40762
                     /rpt_family="MaLR"
     repeat_region   40763. .41007
                     /rpt_family="L2"
     repeat_region   41064. .41231
                     /rpt_family="L2"
     repeat_region   44219. .44279
                     /rpt_family="MaLR"
     repeat_region   44444. .44811
                     /rpt_family="L2"
     repeat_region   44962. .45131
                     /rpt_family="MIR"
     repeat_region   45570. .45729
                     /rpt_family="MIR"
     repeat_region   46229. .46302
                     /rpt_family="MIR"

  Query Match             85.5%;  Score 18.8;  DB 9;  Length 180970;
  Best Local Similarity   90.9%;  Pred. No. 85;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGACTTCTGAGTAACAATCAA 22
           |||||||||||| || ||||||
Db   30424 AAGACTTCTGAGCAAAAATCAA 30445


RESULT 6
AC067929
LOCUS       AC067929              182152 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 6 clone RP11-403I20 map 6, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC067929
VERSION     AC067929.2  GI:8247824
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182152)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 6, clone RP11-403I20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182152)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 182152)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 4, 2000 this sequence version replaced gi:7655991.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10103
                Center clone name: 403_I_20
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 173395 bases at least Q40
                Consensus quality: 177417 bases at least Q30
                Consensus quality: 179105 bases at least Q20
                Insert size: 179000; agarose-fp
                Insert size: 180452; sum-of-contigs
                Quality coverage: 4.8 in Q20 bases; agarose-fp
                Quality coverage: 4.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2222: contig of 2222 bp in length
            *    2223 2322: gap of      100 bp
            *    2323     3785: contig of 1463 bp in length
            *    3786 3885: gap of      100 bp
            *    3886     6753: contig of 2868 bp in length
```

```
            *     6754 6853: gap of      100 bp
            *     6854    10036: contig of 3183 bp in length
            *    10037 10136: gap of      100 bp
            *    10137    14537: contig of 4401 bp in length
            *    14538 14637: gap of      100 bp
            *    14638    21177: contig of 6540 bp in length
            *    21178 21277: gap of      100 bp
            *    21278    30240: contig of 8963 bp in length
            *    30241 30340: gap of      100 bp
            *    30341    40123: contig of 9783 bp in length
            *    40124 40223: gap of      100 bp
            *    40224    47999: contig of 7776 bp in length
            *    48000 48099: gap of      100 bp
            *    48100    58235: contig of 10136 bp in length
            *    58236 58335: gap of      100 bp
            *    58336    67041: contig of 8706 bp in length
            *    67042 67141: gap of      100 bp
            *    67142    75921: contig of 8780 bp in length
            *    75922 76021: gap of      100 bp
            *    76022    86527: contig of 10506 bp in length
            *    86528 86627: gap of      100 bp
            *    86628   101334: contig of 14707 bp in length
            *   101335 101434: gap of      100 bp
            *   101435   117630: contig of 16196 bp in length
            *   117631 117730: gap of      100 bp
            *   117731   138293: contig of 20563 bp in length
            *   138294 138393: gap of      100 bp
            *   138394   157841: contig of 19448 bp in length
            *   157842 157941: gap of      100 bp
            *   157942   182152: contig of 24211 bp in length.
FEATURES             Location/Qualifiers
     source          1. .182152
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6"
                     /clone="RP11-403I20"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .2222
                     /note="assembly_fragment"
     misc_feature    2323. .3785
                     /note="assembly_fragment"
     misc_feature    3886. .6753
                     /note="assembly_fragment"
     misc_feature    6854. .10036
                     /note="assembly_fragment"
     misc_feature    10137. .14537
                     /note="assembly_fragment"
     misc_feature    14638. .21177
                     /note="assembly_fragment"
     misc_feature    21278. .30240
                     /note="assembly_fragment"
     misc_feature    30341. .40123
                     /note="assembly_fragment"
     misc_feature    40224. .47999
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    48100. .58235
                     /note="assembly_fragment"
     misc_feature    58336. .67041
                     /note="assembly_fragment"
     misc_feature    67142. .75921
                     /note="assembly_fragment"
     misc_feature    76022. .86527
                     /note="assembly_fragment"
     misc_feature    86628. .101334
                     /note="assembly_fragment"
     misc_feature    101435. .117630
                     /note="assembly_fragment"
     misc_feature    117731. .138293
                     /note="assembly_fragment"
     misc_feature    138394. .157841
                     /note="assembly_fragment"
     misc_feature    157942. .182152
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    51611 a  40355 c  40613 g  47868 t   1705 others
ORIGIN

  Query Match             85.5%;  Score 18.8;  DB 2;  Length 182152;
  Best Local Similarity   90.9%;  Pred. No. 85;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGACTTCTGAGTAACAATCAA 22
           || ||||||||||||||| |||
Db 152828 AACACTTCTGAGTAACAAACAA 152849


RESULT 7
AC012458
LOCUS       AC012458              189281 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 15 clone RP11-522B15, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC012458
VERSION     AC012458.4  GI:8570271
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 189281)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189281)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 17, 2000 this sequence version replaced gi:6850536.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0522B15
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 77%
            Sequencing vector: plasmid; 23%
            Chemistry: Dye-primer ET; 77% of reads
            Chemistry: Dye-terminator Big Dye; 23% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 183385 bases at least Q40
            Consensus quality: 185502 bases at least Q30
            Consensus quality: 186573 bases at least Q20
            Insert size: 197000; agarose-fp
            Insert size: 188381; sum-of-contigs
            Quality coverage: 4.50 in Q20 bases; agarose-fp
            Quality coverage: 4.71 in Q20 bases; sum-of-contigs
            ----------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     4603: contig of 4603 bp in length
            *    4604     4703: gap of unknown length
            *    4704     9761: contig of 5058 bp in length
            *    9762     9861: gap of unknown length
```

```
            *     9862    18746: contig of 8885 bp in length
            *    18747    18846: gap of unknown length
            *    18847    27396: contig of 8550 bp in length
            *    27397    27496: gap of unknown length
            *    27497    35489: contig of 7993 bp in length
            *    35490    35589: gap of unknown length
            *    35590    51472: contig of 15883 bp in length
            *    51473    51572: gap of unknown length
            *    51573    65229: contig of 13657 bp in length
            *    65230    65329: gap of unknown length
            *    65330    89666: contig of 24337 bp in length
            *    89667    89766: gap of unknown length
            *    89767   117943: contig of 28177 bp in length
            *   117944   118043: gap of unknown length
            *   118044   189281: contig of 71238 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189281
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /clone="RP11-522B15"
     misc_feature    1. .4603
                     /note="assembly_name:Contig6"
     misc_feature    4704. .9761
                     /note="assembly_name:Contig7"
     misc_feature    9862. .18746
                     /note="assembly_name:Contig8"
     misc_feature    18847. .27396
                     /note="assembly_name:Contig9"
     misc_feature    27497. .35489
                     /note="assembly_name:Contig10"
     misc_feature    35590. .51472
                     /note="assembly_name:Contig11"
     misc_feature    51573. .65229
                     /note="assembly_name:Contig12
                     clone_end:SP6
                     vector_side:right"
     misc_feature    65330. .89666
                     /note="assembly_name:Contig13"
     misc_feature    89767. .117943
                     /note="assembly_name:Contig14
                     clone_end:T7
                     vector_side:left"
     misc_feature    118044. .189281
                     /note="assembly_name:Contig15"
BASE COUNT    54120 a  39374 c  39106 g  55769 t    912 others
ORIGIN

  Query Match             85.5%;  Score 18.8;  DB 2;  Length 189281;
  Best Local Similarity   90.9%;  Pred. No. 85;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 AAGACTTCTGAGTAACAATCAA 22
          |||||||||| ||||||| ||
Db 181216 AAGACTTCTGATTAACAATGAA 181237


RESULT 8
AC087477
LOCUS       AC087477               191434 bp    DNA     linear   PRI 16-JUL-2002
DEFINITION  Homo sapiens chromosome 15, clone RP11-522B15, complete sequence.
ACCESSION   AC087477
VERSION     AC087477.8  GI:21844630
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191434)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-522B15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191434)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 191434)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 191434)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 16, 2002 this sequence version replaced gi:20043143.
```

```
        All repeats were identified using RepeatMasker:
        Smit, A.F.A. & Green, P. (1996-1997)
        http://ftp.genome.washington.edu/RM/RepeatMasker.html
        -------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
        -------------- Project Information
            Center project name: L11935
            Center clone name: 522_B_15
        --------------
FEATURES             Location/Qualifiers
     source          1. .191434
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-522B15"
                     /clone_lib="RPCI-11 Human Male BAC"
     unsure          1. .777
                     /note="PCR product sequence only"
     repeat_region   2518. .2556
                     /rpt_family="MIR"
     repeat_region   3963. .4011
                     /rpt_family="(CA)n"
     repeat_region   4103. .4186
                     /rpt_family="C-rich"
     repeat_region   4898. .4965
                     /rpt_family="CT-rich"
     repeat_region   5152. .5213
                     /rpt_family="G-rich"
     repeat_region   5215. .5241
                     /rpt_family="(GA)n"
     unsure          6304. .6308
                     /note="<30 qual SNGL region"
     unsure          6563. .6569
                     /note="<30 qual SNGL region"
     repeat_region   7379. .7470
                     /rpt_family="CT-rich"
     repeat_region   7515. .7548
                     /rpt_family="(TG)n"
     repeat_region   7549. .7602
                     /rpt_family="(GA)n"
     repeat_region   10343. .10363
                     /rpt_family="(TGGGGG)n"
     unsure          10422. .10428
                     /note="<30 qual SNGL region"
     repeat_region   11032. .11060
                     /rpt_family="GC_rich"
     repeat_region   11925. .11954
                     /rpt_family="(T)n"
     repeat_region   13905. .13926
                     /rpt_family="(TC)n"
     unsure          complement(14459. .14464)
                     /note="<30 qual SNGL region"
     unsure          complement(14475. .14479)
                     /note="<30 qual SNGL region"
     unsure          complement(14493. .14498)
                     /note="<30 qual SNGL region"
     repeat_region   18818. .19134
                     /rpt_family="AluSq"
     repeat_region   complement(20563. .20758)
                     /rpt_family="MIR"
     repeat_region   complement(20896. .21016)
                     /rpt_family="MIR"
     repeat_region   21079. .21098
                     /rpt_family="(CA)n"
     repeat_region   21150. .21200
                     /rpt_family="(TATATG)n"
     repeat_region   complement(21400. .21579)
                     /rpt_family="MIR"
     repeat_region   21586. .21876
                     /rpt_family="AluSx"
     repeat_region   complement(22988. .23113)
                     /rpt_family="L2"
     repeat_region   complement(23359. .23666)
                     /rpt_family="AluSq"
     repeat_region   complement(23784. .24062)
                     /rpt_family="AluJo"
     repeat_region   24327. .24350
                     /rpt_family="(A)n"
     repeat_region   24471. .24597
                     /rpt_family="L2"
     repeat_region   25115. .25363
                     /rpt_family="AluJb"
     repeat_region   complement(25426. .25712)
                     /rpt_family="AluSc"
     repeat_region   26549. .26667
                     /rpt_family="FLAM_A"
     repeat_region   complement(26714. .27135)
                     /rpt_family="AluJo"
     repeat_region   27835. .27885
                     /rpt_family="AT_rich"
     repeat_region   complement(29173. .29331)
                     /rpt_family="AluJo"
     repeat_region   complement(29332. .30777)
                     /rpt_family="L1PA5"
     repeat_region   complement(30778. .30790)
                     /rpt_family="AluJo"
     repeat_region   complement(30791. .31103)
                     /rpt_family="AluSp"
     repeat_region   complement(31104. .31235)
                     /rpt_family="AluJo"
     repeat_region   31242. .31310
                     /rpt_family="MLT1F1"
     repeat_region   31437. .31741
                     /rpt_family="MLT1F1"
     repeat_region   31760. .31821
                     /rpt_family="L2"
     repeat_region   31829. .31851
                     /rpt_family="AT_rich"
     repeat_region   complement(31951. .31993)

  Query Match             85.5%;  Score 18.8;  DB 9;  Length 191434;
  Best Local Similarity   90.9%;  Pred. No. 85;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGACTTCTGAGTAACAATCAA 22
           |||||||||||  ||||||| ||
Db 137915 AAGACTTCTGATTAACAATGAA 137936


RESULT 9
AL772303/c
LOCUS       AL772303              197437 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Mus musculus chromosome 2 clone RP23-185P20, *** SEQUENCING IN
            PROGRESS ***, 6 unordered pieces.
ACCESSION   AL772303
VERSION     AL772303.3  GI:21748303
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 197437)
  AUTHORS   Almeida,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 14, 2002 this sequence version replaced gi:21540206.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
```

```
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bM185P20
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 196057 bases at least Q40
            Consensus quality: 196481 bases at least Q30
            Consensus quality: 196768 bases at least Q20
            Insert size: 196937; sum-of-contigs
            Insert size: 198365; 3.5% error; agarose-fp
            Quality coverage: 5.88x in Q20 bases; sum-of-contigs Quality
            coverage: 5.99x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    46842: contig of 46842 bp in length
            *    46843 46942: gap of      100 bp
            *    46943    51200: contig of 4258 bp in length
            *    51201 51300: gap of      100 bp
            *    51301    62621: contig of 11321 bp in length
            *    62622 62721: gap of      100 bp
            *    62722    92981: contig of 30260 bp in length
            *    92982 93081: gap of      100 bp
            *    93082   143175: contig of 50094 bp in length
            *   143176 143275: gap of      100 bp
            *   143276   197437: contig of 54162 bp in length.
FEATURES             Location/Qualifiers
     source          1. .197437
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-185P20"
                     /clone_lib="RPCI-23"
     misc_feature    1. .46842
                     /note="assembly_fragment:00188
                     fragment_chain:1"
     misc_feature    46943. .51200
                     /note="assembly_fragment:01482
                     fragment_chain:1"
     misc_feature    51301. .62621
                     /note="assembly_fragment:01781
                     fragment_chain:1"
     misc_feature    62722. .92981
                     /note="assembly_fragment:00220
                     fragment_chain:2"
     misc_feature    93082. .143175
                     /note="assembly_fragment:00782
                     fragment_chain:2"
     misc_feature    143276. .197437
                     /note="assembly_fragment:00847"
BASE COUNT    59973 a  40264 c  38730 g  57970 t    500 others
ORIGIN

  Query Match             85.5%;  Score 18.8;  DB 2;  Length 197437;
  Best Local Similarity   90.9%;  Pred. No. 85;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 AAGACTTCTGAGTAACAATCAA 22
          |||||||||  | ||||||||||
Db 189362 AAGACTTCTAATTAACAATCAA 189341


RESULT 10
AC127239
LOCUS       AC127239              223761 bp    DNA     linear   HTG 17-JUL-2002
DEFINITION  Mus musculus chromosome UNK clone RP24-381A23, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC127239
VERSION     AC127239.1  GI:21747652
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 223761)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 223761)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 223761)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information ----------
            Center project name: M_BB0381A23
            ----------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 221262 bases at least Q40
            Consensus quality: 221869 bases at least Q30
            Consensus quality: 222233 bases at least Q20
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6570: contig of 6570 bp in length
            *     6571     6670: gap of unknown length
            *     6671    22737: contig of 16067 bp in length
            *    22738    22837: gap of unknown length
            *    22838    48321: contig of 25484 bp in length
            *    48322    48421: gap of unknown length
            *    48422   118533: contig of 70112 bp in length
            *   118534   118633: gap of unknown length
            *   118634   212339: contig of 93706 bp in length
            *   212340   212439: gap of unknown length
            *   212440   216233: contig of 3794 bp in length
            *   216234   216333: gap of unknown length
            *   216334   223761: contig of 7428 bp in length.
FEATURES             Location/Qualifiers
     source          1. .223761
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP24-381A23"
     misc_feature    1. .6570
                     /note="assembly_name:Contig10"
     misc_feature    6671. .22737
                     /note="assembly_name:Contig11"
     misc_feature    22838. .48321
```

```
                 /note="assembly_name:Contig12"
     misc_feature    48422. .118533
                 /note="assembly_name:Contig13"
     misc_feature    118634. .212339
                 /note="assembly_name:Contig14"
     misc_feature    212440. .216233
                 /note="assembly_name:Contig8"
     misc_feature    216334. .223761
                 /note="assembly_name:Contig9"
BASE COUNT    71420 a  41598 c  40871 g  69263 t    609 others
ORIGIN

  Query Match             85.5%;  Score 18.8;  DB 2;  Length 223761;
  Best Local Similarity   90.9%;  Pred. No. 84;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 AAGACTTCTGAGTAACAATCAA 22
          |||||||| ||| ||||||||
Db  87767 AAGACTTCTTAGTTACAATCAA 87788


RESULT 11
AC098888
LOCUS       AC098888              229247 bp    DNA     linear   ROD 21-JUN-2002
DEFINITION  Mus musculus clone RP23-122N8, complete sequence.
ACCESSION   AC098888
VERSION     AC098888.4  GI:19909498
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 229247)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 229247)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2001) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 229247)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 229247)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Apr 3, 2002 this sequence version replaced gi:17105321.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: M_BA0122N08
            ------------------------------------------------------.
FEATURES             Location/Qualifiers
     source          1. .229247
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-122N8"
BASE COUNT    70714 a  42415 c  42708 g  73410 t
ORIGIN

  Query Match             85.5%;  Score 18.8;  DB 10;  Length 229247;
  Best Local Similarity   90.9%;  Pred. No. 83;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 AAGACTTCTGAGTAACAATCAA 22
          |||||||| ||| ||||||||
Db  83202 AAGACTTCTTAGTTACAATCAA 83223


RESULT 12
AC099415
LOCUS       AC099415              317515 bp    DNA     linear   HTG 05-JUN-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-122D8, WORKING DRAFT
            SEQUENCE, 56 unordered pieces.
ACCESSION   AC099415
VERSION     AC099415.3  GI:21326409
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 317515)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 317515)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2001) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 317515)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Jun 5, 2002 this sequence version replaced gi:20069750.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: M_BA0122D08
            ------------------------------------------------------
            ---------------- Summary Statistics  ----------------
            Sequencing vector: M13; 16%
            Sequencing vector: plasmid; 84%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 293243 bases at least Q40
            Consensus quality: 303369 bases at least Q30
            Consensus quality: 309642 bases at least Q20
            Insert size: 224000; agarose-fp
            Insert size: 322378; sum-of-contigs
            Quality coverage: 37.21 in Q20 bases; agarose-fp
            Quality coverage: 24.94 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 56 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1211: contig of 1211 bp in length
            *    1212     1311: gap of unknown length
            *    1312     2357: contig of 1046 bp in length
            *    2358     2457: gap of unknown length
            *    2458     3574: contig of 1117 bp in length
            *    3575     3674: gap of unknown length
            *    3675     4817: contig of 1143 bp in length
            *    4818     4917: gap of unknown length
            *    4918     6015: contig of 1098 bp in length
            *    6016     6115: gap of unknown length
```

```
*     6116     7259: contig of 1144 bp in length
*     7260     7359: gap of unknown length
*     7360     8831: contig of 1472 bp in length
*     8832     8931: gap of unknown length
*     8932    10179: contig of 1248 bp in length
*    10180    10279: gap of unknown length
*    10280    11586: contig of 1307 bp in length
*    11587    11686: gap of unknown length
*    11687    12940: contig of 1254 bp in length
*    12941    13040: gap of unknown length
*    13041    14154: contig of 1114 bp in length
*    14155    14254: gap of unknown length
*    14255    15728: contig of 1474 bp in length
*    15729    15828: gap of unknown length
*    15829    17212: contig of 1384 bp in length
*    17213    17312: gap of unknown length
*    17313    18620: contig of 1308 bp in length
*    18621    18720: gap of unknown length
*    18721    20283: contig of 1563 bp in length
*    20284    20383: gap of unknown length
*    20384    21676: contig of 1293 bp in length
*    21677    21776: gap of unknown length
*    21777    23147: contig of 1371 bp in length
*    23148    23247: gap of unknown length
*    23248    24781: contig of 1534 bp in length
*    24782    24881: gap of unknown length
*    24882    26627: contig of 1746 bp in length
*    26628    26727: gap of unknown length
*    26728    28451: contig of 1724 bp in length
*    28452    28551: gap of unknown length
*    28552    29863: contig of 1312 bp in length
*    29864    29963: gap of unknown length
*    29964    31941: contig of 1978 bp in length
*    31942    32041: gap of unknown length
*    32042    33660: contig of 1619 bp in length
*    33661    33760: gap of unknown length
*    33761    34889: contig of 1129 bp in length
*    34890    34989: gap of unknown length
*    34990    36451: contig of 1462 bp in length
*    36452    36551: gap of unknown length
*    36552    38157: contig of 1606 bp in length
*    38158    38257: gap of unknown length
*    38258    40052: contig of 1795 bp in length
*    40053    40152: gap of unknown length
*    40153    41725: contig of 1573 bp in length
*    41726    41825: gap of unknown length
*    41826    42996: contig of 1171 bp in length
*    42997    43096: gap of unknown length
*    43097    44639: contig of 1543 bp in length
*    44640    44739: gap of unknown length
*    44740    46047: contig of 1308 bp in length
*    46048    46147: gap of unknown length
*    46148    47831: contig of 1684 bp in length
*    47832    47931: gap of unknown length
*    47932    49690: contig of 1759 bp in length
*    49691    49790: gap of unknown length
*    49791    51094: contig of 1304 bp in length
*    51095    51194: gap of unknown length
*    51195    53501: contig of 2307 bp in length
*    53502    53601: gap of unknown length
*    53602    55272: contig of 1671 bp in length
*    55273    55372: gap of unknown length
*    55373    57428: contig of 2056 bp in length
*    57429    57528: gap of unknown length
*    57529    59819: contig of 2291 bp in length
*    59820    59919: gap of unknown length
*    59920    61717: contig of 1798 bp in length
*    61718    61817: gap of unknown length
*    61818    64311: contig of 2494 bp in length
*    64312    64411: gap of unknown length
*    64412    67042: contig of 2631 bp in length
*    67043    67142: gap of unknown length
*    67143    69238: contig of 2096 bp in length
*    69239    69338: gap of unknown length
*    69339    71520: contig of 2182 bp in length
*    71521    71620: gap of unknown length
*    71621    74319: contig of 2699 bp in length
*    74320    74419: gap of unknown length
*    74420    76137: contig of 1718 bp in length
*    76138    76237: gap of unknown length
*    76238    79611: contig of 3374 bp in length
*    79612    79711: gap of unknown length
*    79712    82702: contig of 2991 bp in length
*    82703    82802: gap of unknown length
*    82803    84448: contig of 1646 bp in length
*    84449    84548: gap of unknown length
*    84549    87422: contig of 2874 bp in length
*    87423    87522: gap of unknown length
*    87523    89088: contig of 1566 bp in length
*    89089    89188: gap of unknown length
*    89189    92189: contig of 3001 bp in length
*    92190    92289: gap of unknown length
*    92290    96604: contig of 4315 bp in length
*    96605    96704: gap of unknown length
*    96705    98834: contig of 2130 bp in length
*    98835    98934: gap of unknown length
*    98935   109620: contig of 10686 bp in length
*   109621   109720: gap of unknown length
*   109721   177250: contig of 67530 bp in length
*   177251   177350: gap of unknown length
*   177351   317515: contig of 140165 bp in length.
FEATURES             Location/Qualifiers
     source          1. .317515
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-122D8"
     misc_feature    1. .1211
                     /note="assembly_name:Contig89"
     misc_feature    1312. .2357
                     /note="assembly_name:Contig93"
     misc_feature    2458. .3574
                     /note="assembly_name:Contig141"
     misc_feature    3675. .4817
                     /note="assembly_name:Contig142"
     misc_feature    4918. .6015
                     /note="assembly_name:Contig147"
     misc_feature    6116. .7259
                     /note="assembly_name:Contig159"
     misc_feature    7360. .8831
                     /note="assembly_name:Contig163"
     misc_feature    8932. .10179
                     /note="assembly_name:Contig165"
     misc_feature    10280. .11586
                     /note="assembly_name:Contig170"
     misc_feature    11687. .12940
                     /note="assembly_name:Contig174"
     misc_feature    13041. .14154
                     /note="assembly_name:Contig177"
     misc_feature    14255. .15728
                     /note="assembly_name:Contig179"
     misc_feature    15829. .17212
                     /note="assembly_name:Contig180"
     misc_feature    17313. .18620

  Query Match             85.5%;  Score 18.8;  DB 2;  Length 317515;
  Best Local Similarity   90.9%;  Pred. No. 81;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGACTTCTGAGTAACAATCAA 22
           ||||||||| ||| ||||||||
Db   17367 AAGACTTCTTAGTTACAATCAA 17388


RESULT 13
AE011043/c
```

```
LOCUS       AE011043              10243 bp    DNA     linear   BCT 03-APR-2002
DEFINITION  Methanosarcina acetivorans str. C2A, section 388 of 534 of the
            complete genome.
ACCESSION   AE011043 AE010299
VERSION     AE011043.1  GI:19917413
KEYWORDS    .
SOURCE      Methanosarcina acetivorans C2A.
  ORGANISM  Methanosarcina acetivorans C2A
            Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
            Methanosarcinaceae; Methanosarcina.
REFERENCE   1  (bases 1 to 10243)
  AUTHORS   Galagan,J.E., Nusbaum,C., Roy,A., Endrizzi,M.G., Macdonald,P.,
            FitzHugh,W., Calvo,S., Engels,R., Smirnov,S., Atnoor,D., Brown,A.,
            Allen,N., Naylor,J., Stange-Thomann,N., DeArellano,K., Johnson,R.,
            Linton,L., McEwan,P., McKernan,K., Talamas,J., Tirrell,A., Ye,W.,
            Zimmer,A., Barber,R.D., Cann,I., Graham,D.E., Grahame,D.A.,
            Guss,A., Hedderich,R., Ingram-Smith,C., Kuettner,C.H.,
            Krzycki,J.A., Leigh,J.A., Li,W., Liu,J., Mukhopadhyay,B.,
            Reeve,J.N., Smith,K., Springer,T.A., Umayam,L.A., White,O.,
            White,R.H., de Macario,E.C., Ferry,J.G., Jarrell,K.F., Jing,H.,
            Macario,A.J.L., Paulsen,I., Pritchett,M., Sowers,K.R.,
            Swanson,R.V., Zinder,S.H., Lander,E., Metcalf,W.W. and Birren,B.
  TITLE     The Genome of M. acetivorans Reveals Extensive Metabolic and
            Physiological Diversity
  JOURNAL   Genome Res. 12 (4), 532-542 (2002)
  MEDLINE   21929760
   PUBMED   11932238
REFERENCE   2  (bases 1 to 10243)
  AUTHORS   Birren,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2002) Center for Genome Research, Whitehead
            Institute, Nine Cambridge Center, Cambridge, MA 02141, USA
FEATURES             Location/Qualifiers
     source          1. .10243
                     /organism="Methanosarcina acetivorans C2A"
                     /strain="C2A"
                     /db_xref="taxon:188937"
     gene            complement(88. .603)
                     /gene="MA3371"
     CDS             complement(88. .603)
                     /gene="MA3371"
                     /codon_start=1
                     /transl_table=11
                     /product="predicted protein"
                     /protein_id="AAM06740.1"
                     /db_xref="GI:19917414"
                     /translation="MIRESNNIMAIMGGRETGRGPLQLDIGIVVSGLGGNSNQERNQF
                     NGLFSFSGIDFGLYIGPSYNPFQTIGPSIGIGLVIGSNGFLAAKLGIHFRRAFGLGML
                     FPIVPVYGVSGASFLAALPVSEGAKAAATIPLLMVPAAVVYQENIKELRRQGTEILMK
                     LRSYISGKRTF"
     gene            complement(859. .1488)
                     /gene="MA3372"
     CDS             complement(859. .1488)
                     /gene="MA3372"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAM06741.1"
                     /db_xref="GI:19917415"
                     /translation="MKIAVVENENQQTSSIFEPGFIAVYEEDGVEWKVLNRFENQVCN
                     ARGMAAVRMAVGDAIKQLGDVKVVVASEIPGIASGTFQAAGFDIFLVDSKVLDVLDLI
                     KKDMLETIEKRQEEPRKFDITEFLEPGENRGDFSINIEDIMFKFPDLTSKKILIPYLK
                     NGEFNRLDVICGHMPKWFVTDLGAMGFEYETVNESTNRKTVRVVRAQTP"
     gene            1981. .3642
                     /gene="ilvD"
                     /note="MA3373"
     CDS             1981. .3642
                     /gene="ilvD"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroxy-acid dehydratase"
                     /protein_id="AAM06742.1"
                     /db_xref="GI:19917416"
                     /translation="MRSDIIKEGPERAPNRSLLKATGVTDSEMKKPFIAVVNSWNDII
                     PGHMHLNKLAEAVKAGIRNAGGVPFEFHTVGVCDGIAMGHEGMKYSLPSREIIEDTIE
                     LMVKAHQFDGMVLIPSCDKIVPGHLMAAGRLDIPAIVVTGGPMLPGYVDDNYRDLISV
                     SEGVGAYSAGKISETEFKRLENLSCAGAGSCAGMFTANTMACMTEALGLSLPGCATAH
                     AVDAKKILIAKESGERIVELVKENLTPRKIVTQKSFENAIMVDMAVGGSTNTTLHLPA
                     LAHEFGLGLPLEAFDELSRKTPHLISLRPGGPNFILHFDRAGGVEAVMQRLASKLHLD
                     QLTVNGKTIGENLDELEIVNPKLNAEIITTLENPIHAEGGIAVLKGSLAPDGSVVKQA
                     AVDPKMHVYTGPAKVYDCEEDAMKSILAGDVKPGDIVVIRYEGPKGGPGMREMLAATA
                     ALGGRGLLESVALVTDGRFSGGTRGPCIGHVSPEASDGGPIALVKDGDLIEINIPERA
                     LNLKISEEELKQRKASFVPPKTEITGYLARYQRAVHSANTGGIVD"
     gene            complement(3805. .4149)
                     /gene="MA3374"
     CDS             complement(3805. .4149)
                     /gene="MA3374"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAM06743.1"
                     /db_xref="GI:19917417"
                     /translation="MWIAVDRVGKKFINCSFGSRGTKTGQLLWEKLKKKEIGEVMTDH
                     WRAYAEFIPETIHTQSKAETYTVEGYNGILRHFLARLRRKTKCYTKSLEMLKYSVLLL
                     MKHRNKELSIFN"
     gene            complement(4053. .4487)
                     /gene="MA3375"
     CDS             complement(4053. .4487)
                     /gene="MA3375"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAM06744.1"
                     /db_xref="GI:19917418"
                     /translation="MNCPRCKSSNHTKNGIVCGRQRYKCHDCGYNYSVELKSTASSPL
                     VKRQALQLYLEGLGFRSIGRFLGVSHVSVQKWIKKFGQEIEELKSENEISIVELDEMH
                     TYIGNKKNIAGSGLLLIELGKNSSTALLVAEERKLDNYSGKN"
     gene            4675. .4950
                     /gene="MA3376"
     CDS             4675. .4950
                     /gene="MA3376"
                     /codon_start=1
                     /transl_table=11
                     /product="predicted protein"
                     /protein_id="AAM06745.1"
                     /db_xref="GI:19917419"
                     /translation="MGEVIATRLFGGELIKRSQSDVTEFITGRGIGYPATKRDIIRFA
                     TGKSVESDVLDLLKGIPEIEYNTPDEVTREIERLESQRTREYTRPEY"
     gene            5669. .7393
                     /gene="MA3377"
     CDS             5669. .7393
                     /gene="MA3377"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAM06746.1"
                     /db_xref="GI:19917420"
                     /translation="MCTTLIITRGASKDGSMIVAHSDDDELGDQRIIYVPAQDHEQGA
                     MRQVFEEHYRYPRLVTYERGPNYEPATHHVSQSPNPVPDLTGPIGYIPQVRHTYAYFD
                     GNYGIMNEHNLMMGECTNGAKYQPAHVSAEEAEKNSKHIRLFYSQELSRVALERCNKA
                     REAVELMGKLLDKYGYYSTGETLLVADEDEAWVFEMCALPDEQYHSAWVAQQVPDGEV
                     FVAANEFRIRQIEDNRRDIICSELLKPGLKNLGWWNPEDGPLDWLSSVSNGEYNHPYY
                     SLRRVWRVLDRVNPDLALSPWVTRGFKDVKDGSYAIDYPFSIKPKCPLTLQDVFALYR
                     DHYEGTQFDLTRGVAAGPYGDPNRFVGPYDGNQNNVTENEHMYGAWERPISIFYQGYT
                     YVNQLRPKTPEIPEASKGLCWFGPDVSYTTCFVPFAAQALELPERYQSGDPQVFDPKT
                     AWWAFDFVSNWARLNYQRMTRVDILPLQQKIETEEVQKTILQWDECCRGKSDLEARDL
                     LTRLGKENANEVVKQWHKLGHALIAKYSDGYCNLPENQEPQNIGYSSKWLALTNYKDG
                     PTSYEMKP"
     gene            complement(7624. .7914)
                     /gene="MA3378"
     CDS             complement(7624. .7914)
                     /gene="MA3378"
                     /codon_start=1
                     /transl_table=11
                     /product="predicted protein"
                     /protein_id="AAM06747.1"
```

```
                     /db_xref="GI:19917421"
                     /translation="MISMDSMEKIFGKTAQMTVLKNLIENQNEPTYLSGIAEETGLSH
                     SSVSRVITPLIESGVILEKPLGKQIRTFQLNMESEAAKLIVDFYYKINQLVE"
     gene            complement(8740. .10059)
                     /gene="MA3379"
     CDS             complement(8740. .10059)
                     /gene="MA3379"
                     /codon_start=1
                     /transl_table=11
                     /product="amino acid transporter"
                     /protein_id="AAM06748.1"
                     /db_xref="GI:19917422"
                     /translation="MFEVKYNSGNYMGRESELKRDLSLLEITLVGIGDILGAGIYVLM
                     GKAAGLAGNMVWASFLFAGFTASLSALSYMELSSMYPRAGAEYVFARKAFGERIGLFV
                     GLLVIYYVAITASAVALGFGRYFSNIFGGGILIGAISLYILLSLVMVYGIKESARLAI
                     LMTFIELSGLMIIIYVGIPYLGTVNYFETPSLSGVFEASALIIFAFLGFEDIVRLSQE
                     TKDAEKTTPKALLISIFFTIFLYICVAVASVSVLDFQVLGISEIPLAEVAAVAFGEKA
                     FVLLSWIALFSTMNTVLVVMLGGSRIVYGMANSGSLPKILARVHQKLKTPWTAICGIA
                     FFSSLFVFLGDIATVANIANFMIFIVFFIVNISLIKLRYTDPERKRPFRVPVSIGRFP
                     IFPALGALSAVFLFSQIGKEVMLYGFFLAGISALLVLLKTRKRENEL"
BASE COUNT     3037 a   2258 c   2111 g   2837 t
ORIGIN

  Query Match             83.6%;  Score 18.4;  DB 1;  Length 10243;
  Best Local Similarity   95.0%;  Pred. No. 1.8e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AAGACTTCTGAGTAACAATC 20
           | ||||||||||||||||||
Db    2740 AGGACTTCTGAGTAACAATC 2721


RESULT 14
AE010860/c
LOCUS       AE010860               10467 bp    DNA     linear   BCT 03-APR-2002
DEFINITION  Methanosarcina acetivorans str. C2A, section 205 of 534 of the
            complete genome.
ACCESSION   AE010860 AE010299
VERSION     AE010860.1  GI:19915697
KEYWORDS    .
SOURCE      Methanosarcina acetivorans C2A.
  ORGANISM  Methanosarcina acetivorans C2A
            Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
            Methanosarcinaceae; Methanosarcina.
REFERENCE   1  (bases 1 to 10467)
  AUTHORS   Galagan,J.E., Nusbaum,C., Roy,A., Endrizzi,M.G., Macdonald,P.,
            FitzHugh,W., Calvo,S., Engels,R., Smirnov,S., Atnoor,D., Brown,A.,
            Allen,N., Naylor,J., Stange-Thomann,N., DeArellano,K., Johnson,R.,
            Linton,L., McEwan,P., McKernan,K., Talamas,J., Tirrell,A., Ye,W.,
            Zimmer,A., Barber,R.D., Cann,I., Graham,D.E., Grahame,D.A.,
            Guss,A., Hedderich,R., Ingram-Smith,C., Kuettner,C.H.,
            Krzycki,J.A., Leigh,J.A., Li,W., Liu,J., Mukhopadhyay,B.,
            Reeve,J.N., Smith,K., Springer,T.A., Umayam,L.A., White,O.,
            White,R.H., de Macario,E.C., Ferry,J.G., Jarrell,K.F., Jing,H.,
            Macario,A.J.L., Paulsen,I., Pritchett,M., Sowers,K.R.,
            Swanson,R.V., Zinder,S.H., Lander,E., Metcalf,W.W. and Birren,B.
  TITLE     The Genome of M. acetivorans Reveals Extensive Metabolic and
            Physiological Diversity
  JOURNAL   Genome Res. 12 (4), 532-542 (2002)
  MEDLINE   21929760
   PUBMED   11932238
REFERENCE   2  (bases 1 to 10467)
  AUTHORS   Birren,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2002) Center for Genome Research, Whitehead
            Institute, Nine Cambridge Center, Cambridge, MA 02141, USA
FEATURES             Location/Qualifiers
     source          1. .10467
                     /organism="Methanosarcina acetivorans C2A"
                     /strain="C2A"
                     /db_xref="taxon:188937"
     gene            complement(341. .1141)
                     /gene="MA1801"
     CDS             complement(341. .1141)
                     /gene="MA1801"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAM05207.1"
                     /db_xref="GI:19915698"
                     /translation="MYCWNPELYAFSSSAQKSWGIELLTKFPLKGNERVLDVGCGDGK
                     LSAEIAKRLPEGSVLGIDLSEAMVCFAKNHYPKEQFPNLSFMLMDAGNVPFESEFDVI
                     FSNAALHWIKEPKAIETVLKGFLKSLRPEGKLLAQFGGRGNAAEVLLVLNSMLEDEKW
                     SPYFGNFVFPYGFYGPEEYGKWLKNAGFSVRRAELYSKDMTLKGKEGLSGWFASTWHP
                     FTQRVPENLKEEFIDELAVNFIKLHPLDNEGYVHIRMVRLEIEADIEK"
     gene            2004. .3665
                     /gene="ilvD"
                     /note="MA1802"
     CDS             2004. .3665
                     /gene="ilvD"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroxy-acid dehydratase"
                     /protein_id="AAM05208.1"
                     /db_xref="GI:19915699"
                     /translation="MRSAIIKEGPERAANRSLLKATGVTDSEMKKPFIAVVNSWNEIV
                     PGHIHLDKLAEAVKAGIRNAGGVPFEFHTIGVCDGLAMGHEGMKYSLPSREVIEDTIE
                     LMVRAHQFDGMVLIPTCDKIVPGHLMAAGRLDIPAIVVTGGPMLPGYVDDKYTDLISV
                     FEGVGAYSAGKLSEAELKRLENLSCAGAGSCAGMFTANTMACMTEALGLSLPGCATAH
                     AVDAKKVRIAKESGERIVALVKENLTPRKIVTQKSFENAIMVDMAVGGSTNTTLHLPA
                     LAHEFGLELPLKTFDELSRTTPHLISLRPGGPNFMLHFDRAGGVEAVVQRLASKLHLD
                     QLTVNGKTIGENLDELEIVNPKLNAEIITTLENPIHAEGGIAVLKGSLAPDGSVVKQA
                     AVDPKMRVHTGPAKVYDCEEDAMKSILAGDVKPGDIVVIRYEGPKGGPGMREMLAATA
                     AIGGMGLLESVALITDGRFSGGTRGPCIGHVSPEASEGGPIGLVKDGDLIEINIPERI
                     LNLKVTEEELEKRKAAFVPPKKEVTGYLARYQRSVHSANTGGIVD"
     gene            4511. .4879
                     /gene="MA1803"
     CDS             4511. .4879
                     /gene="MA1803"
                     /codon_start=1
                     /transl_table=11
                     /product="predicted protein"
                     /protein_id="AAM05209.1"
                     /db_xref="GI:19915700"
                     /translation="MNFSFRFFLIMVKTEFFLIGIFCKHKFLVDSYFCKCKRIIPISH
                     SWQDIQIYMIPDTAFRKKSRNDEPVNSSCIVFAYITAAAWRNLGSFLSYFGPVSPVSH
                     KTNGMVCFILYYSHYTTPLS"
     gene            complement(5796. .6758)
                     /gene="mtxH"
                     /note="MA1804"
     CDS             complement(5796. .6758)
                     /gene="mtxH"
                     /codon_start=1
                     /transl_table=11
                     /product="methyltransferase"
                     /protein_id="AAM05210.1"
                     /db_xref="GI:19915701"
                     /translation="MRSVDFEGVKMFKFQKEQEIANIAGIKVGGQPGELPTLLAGTIF
                     YNKHEIVEDAAKGLFDRIAAEKLVNLQEVGSDMTGNPHMVHIFGTTQESITRYIDFIA
                     EVSESPFLIDSPEGTVRAHASRYVSEIGLADRAVYNSINMSITTSEIEALAQSDIDSS
                     IILGFNAMDSSLRGRMEMLETGAGLLEEGLLSIADRCGIVNKLIDPSITPMGNGAGIA
                     LRMTITAKAKWGYPTGSGIHNAPSAWNWLNRQKEKNPVLYKICDVGSTCLQQAAAGDF
                     ILYGPIEYAQYVFPMAAMSDIMIAEAVADLDIEPASRHPINLLV"
     gene            complement(6986. .7579)
                     /gene="mtxA"
                     /note="MA1805"
     CDS             complement(6986. .7579)
                     /gene="mtxA"
                     /codon_start=1
                     /transl_table=11
                     /product="methyltransferase"
                     /protein_id="AAM05211.1"
                     /db_xref="GI:19915702"
                     /translation="MKTIAENWPPVRGDYVTGKPDSRIAVVTLASHLEAFPNAAIWGS
                     SKTENLGVEKIIVNVISNSNIRYVLVCGTESRGHLAGHSLLAIHANGIDEKGRIVGSQ
                     GAIPFIENISREAVDRFQRQVTLLDRIGLNDPEEIRKIVEDYRDKGEVYPEEPVVVCA
```

```
                     PKKRQASFAVPSSGDVIISEELVMDSMAGIICLAESL"
     gene            complement(7576. .8418)
                     /gene="mtxX"
                     /note="MA1806"
     CDS             complement(7576. .8418)
                     /gene="mtxX"
                     /codon_start=1
                     /transl_table=11
                     /product="methyltransferase"
                     /protein_id="AAM05212.1"
                     /db_xref="GI:19915703"
                     /translation="MYALLEAIEARARMNRARVAMGIRDPNPKTLESAWRAQVLGYAQ
                     VVLVGDKKEIDRIGTDLEVVDTKDPEKVLSELLVSRKVDAAIRGTAKASGTLSNLKEA
                     LGMKRVCRLALLLTADETPFFLAPVGIDEGNTIADKLRMIILGAEHIRRLGVEPVVGV
                     LSGGRIGDLGRDRRVDRTLADGEFVAARALGLGINAKHYTILIEDAIKESNFIVAPDG
                     ISGNLIFRTIAFLGGGDGLGAPVLMDDYVFVDTSRVGGHFTKAIMLASALSYLHKERK
                     KVIR"
     gene            8697. .9032
                     /gene="MA1807"
     CDS             8697. .9032
                     /gene="MA1807"
                     /codon_start=1
                     /transl_table=11
                     /product="predicted protein"
                     /protein_id="AAM05213.1"
                     /db_xref="GI:19915704"
                     /translation="MLENVRIEFLEENARLKHRIWSAELRVENLEKIVHSKIFLNQNR
                     ISGRLIELFEKRDYEAIEEYLSEVLTLETQIEKNGKTVEDIEEIVLAEGGLNGRWLCI
                     IIQYFMPKA"
     gene            9762. .9980
                     /gene="MA1808"
     CDS             9762. .9980
                     /gene="MA1808"
                     /codon_start=1
                     /transl_table=11
                     /product="predicted protein"
                     /protein_id="AAM05214.1"
                     /db_xref="GI:19915705"
                     /translation="MYHAGSYPYLKGQMCDWVSGESESPDRIDTLVWTITELMGKEPH
                     VQAAMTWDGVGVVGRFQNINNMKALIKT"
BASE COUNT     2883 a   2327 c   2252 g   3005 t
ORIGIN

  Query Match             83.6%;  Score 18.4;  DB 1;  Length 10467;
  Best Local Similarity   95.0%;  Pred. No. 1.8e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AAGACTTCTGAGTAACAATC 20
           | ||||||||||||||||||
Db    2763 AGGACTTCTGAGTAACAATC 2744


RESULT 15
CNS01DSY
LOCUS       CNS01DSY              152556 bp    DNA     linear   PRI 06-JUN-2001
DEFINITION  Human chromosome 14 DNA sequence BAC C-2555O16 of library CalTech-D
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL122035
VERSION     AL122035.6  GI:14330253
KEYWORDS    HTG; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152556)
  AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 152556)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Jun 8, 2001 this sequence version replaced gi:12001723.
            --------------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            --------------
            The following BAC sequence is oriented from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : R-973N13 (AC=AL049869)
            Downstream BAC (overlapping the SP6 end) : R-712C19 ----------
            Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 6.31x in Q20 bases; sum-of-contigs
            --------------
            ----------------------
            Overall quality chart :
            Range     :   bases
            0         :
            1  -  9   :
            10 - 19   :      27
            20 - 29   :      97
            30 - 39   :     567
            40 - 49   :    4754
            50 - 59   :    6763
            60 - 69   :    9978
            70 - 79   :   23310
            80 - 89   :   52677
            90 - 99   :   54383
            ----------------------
            Percentage of bases with a quality value >= 40 : 99 %.
FEATURES                Location/Qualifiers
     source             1. .152556
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="14"
                        /clone="C-2555O16"
                        /clone_lib="CalTech-D"
     STS                12290. .12382
                        /note="matching EMBL:Z39911
                        RHdb:RH74596
                        dbSTS:STS51710
                        Identified using the e-PCR software (G. Schuler)"
     STS                16944. .17144
                        /note="matching EMBL:M90359
                        RHdb:RH69194
                        RHdb:RH17748
                        dbSTS:STS14788
                        Identified using the e-PCR software (G. Schuler)"
     STS                26718. .26807
                        /note="matching EMBL:T03579
                        RHdb:RH8964
                        dbSTS:STS1878
                        Identified using the e-PCR software (G. Schuler)"
     STS                38842. .38975
                        /note="matching EMBL:N98589
                        RHdb:RH66227
                        dbSTS:STS46156
                        Identified using the e-PCR software (G. Schuler)"
     STS                144696. .144895
                        /note="matching EMBL:G04328
                        RHdb:RH53667
                        RHdb:RH3777
                        dbSTS:STS26132
                        Identified using the e-PCR software (G. Schuler)"
BASE COUNT     44388 a  32591 c  33034 g  42543 t
ORIGIN

  Query Match             83.6%;  Score 18.4;  DB 9;  Length 152556;
  Best Local Similarity   95.0%;  Pred. No. 1.4e+02;
```

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       3 GACTTCTGAGTAACAATCAA 22
           ||||||||||||| |||||||
Db   29842 GACTTCTGAGTACCAATCAA 29861
```

Search completed: March 17, 2003, 11:38:09
Job time : 574.495 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:42:12 ; Search time 126.086 Seconds
(without alignments)
392.938 Million cell updates/sec

Title: US-09-836-439-6
Perfect score: 22
Sequence: 1 aagacttctgagtaacaatcaa 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18.4	83.6	48908	22	AAK82338	Human immune/haema
	2	17.8	80.9	801	21	AAC42853	Arabidopsis thalia
c	3	17.8	80.9	877	11	AAQ03999	Sequence complemen
	4	17.8	80.9	1353	21	AAC50782	Arabidopsis thalia
c	5	17.2	78.2	513	22	AAK54111	Murine transport a
	6	17.2	78.2	7626	23	ABL07736	Drosophila melanog
	7	17.2	78.2	349980	22	AAH68529	C glutamicum codin
	8	16.8	76.4	461	21	AAC57234	Pinus radiata tran
	9	16.8	76.4	21477	22	AAK66626	Human immune/haema
	10	16.8	76.4	21480	22	AAK66625	Human immune/haema
c	11	16.4	74.5	3114	22	AAS01375	Human TANGO 405 cD
	12	16.4	74.5	29329	22	ABA18026	Human nervous syst
	13	16.4	74.5	29329	22	ABA20511	Human nervous syst
	14	16.4	74.5	29329	22	AAK70791	Human immune/haema
	15	16.4	74.5	29329	22	AAK78512	Human immune/haema
	16	16.4	74.5	640681	24	ABA92787	Buchnera sp. genom
	17	16.2	73.6	318	24	ABN19466	Human ORFX polynuc
c	18	16.2	73.6	354	24	ABL82328	Human ovarian canc
	19	16.2	73.6	531	23	AAS78341	DNA encoding novel
	20	16.2	73.6	531	23	AAS81173	DNA encoding novel
c	21	16.2	73.6	2976	22	AAH17705	Human cDNA sequenc
	22	16.2	73.6	3063	22	AAI58587	Human polynucleoti
c	23	16.2	73.6	3083	17	AAT36036	rchd534 gene diffe
c	24	16.2	73.6	3083	18	AAT94472	Human rchd534 gene
c	25	16.2	73.6	3083	20	AAX26275	cDNA sequence of r
c	26	16.2	73.6	3083	20	AAV81861	Human rchd534 enco
c	27	16.2	73.6	3083	21	AAA88579	Human rchd534 cDNA
c	28	16.2	73.6	3083	21	AAZ89823	Human cardiovascul
c	29	16.2	73.6	3083	21	AAZ88036	Rchd534 gene seque
c	30	16.2	73.6	3084	21	AAZ50712	Nucleotide sequenc
c	31	16.2	73.6	3097	22	AAI60373	Human polynucleoti
	32	16.2	73.6	3199	23	AAS87511	DNA encoding novel
c	33	16.2	73.6	3499	23	AAS78395	DNA encoding novel
	34	16.2	73.6	4732	22	AAK70267	Human immune/haema
c	35	16.2	73.6	4963	23	AAS78396	DNA encoding novel
	36	16.2	73.6	6193	23	AAS87510	DNA encoding novel
	37	16.2	73.6	6305	16	AAQ96127	Tobacco leaf antif
	38	16.2	73.6	9630	23	AAS87490	DNA encoding novel
	39	16.2	73.6	11742	23	AAS87514	DNA encoding novel
	40	16.2	73.6	12027	23	AAS87478	DNA encoding novel
c	41	16.2	73.6	12776	23	ABL11894	Drosophila melanog
	42	16.2	73.6	17727	23	AAS87482	DNA encoding novel
	43	16.2	73.6	19861	23	AAS87516	DNA encoding novel
	44	15.8	71.8	711	18	AAT67444	H. pylori secreted
	45	15.8	71.8	723	18	AAT67946	H. pylori secreted

ALIGNMENTS

RESULT 1
AAK82338
ID AAK82338 standard; DNA; 48908 BP.
XX
AC AAK82338;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37150.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 37150; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to

```
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 48908 BP; 14764 A; 10631 C; 10740 G; 12773 T; 0 other;

  Query Match             83.6%;  Score 18.4;  DB 22;  Length 48908;
  Best Local Similarity   95.0%;  Pred. No. 42;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 GACTTCTGAGTAACAATCAA 22
           |||||||||||| |||||||
Db    3233 GACTTCTGAGTACCAATCAA 3252


RESULT 2
AAC42853
ID   AAC42853 standard; DNA; 801 BP.
XX
AC   AAC42853;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 37100.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
```

PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 80.9%; Score 17.8; DB 21; Length 801;
Best Local Similarity 90.5%; Pred. No. 56;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGACTTCTGAGTAACAATCA 21
||||||||| |||| |||||||
Db 543 AAGACTTCAGAGTCACAATCA 563

RESULT 3
AAQ03999/c
ID AAQ03999 standard; DNA; 877 BP.
XX
AC AAQ03999;
XX
DT 03-SEP-1990 (first entry)
XX
DE Sequence complementary to dystrophin gene.
XX
KW X-chromosome; ornithine transcarbamylase deficiency;
KW muscular dystrophy; dystrophin; ds.
XX
OS Synthetic.
XX
PN EP364255-A.
XX
PD 18-APR-1990.
XX
PF 11-OCT-1989; 89EP-0310424.
XX
PR 12-OCT-1988; 88US-0256689.
XX
PA (BAYU) BAYLOR UNIV COLLEGE.
XX
PI CasKey CT, Chamberlain JS, Gibbs RAL, Rainer JE, Nguyen PN;
XX
DR WPI; 1990-117752/16.
XX
PT Multiplex genomic DNA amplification for deletion detection -
PT useful for detecting X-linked diseases such as ornithine
PT transcarbamylase deficiency and X-linked muscular dystrophy.
XX
PS Claim 9; Page 19; 32pp; English.
XX
CC Paired oligonucleotide primers are used in detecting deletions
CC specifically of the X and Y chromosomes.
CC Dystrophin gene may be isolated this way.
XX
SQ Sequence 877 BP; 289 A; 160 C; 131 G; 297 T; 0 other;

Query Match 80.9%; Score 17.8; DB 11; Length 877;
Best Local Similarity 90.5%; Pred. No. 56;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGACTTCTGAGTAACAATCA 21
||| ||||| ||||||||||
Db 132 AAGCCTTCTCAGTAACAATCA 112

RESULT 4
AAC50782
ID AAC50782 standard; DNA; 1353 BP.
XX
AC AAC50782;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 66116.

XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

```
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             80.9%;  Score 17.8;  DB 21;  Length 1353;
  Best Local Similarity   90.5%;  Pred. No. 59;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AAGACTTCTGAGTAACAATCA 21
           |||||||| |||| |||||||
Db     543 AAGACTTCAGAGTCACAATCA 563


RESULT 5
AAK54111/c
ID   AAK54111 standard; cDNA; 513 BP.
XX
AC   AAK54111;
XX
DT   16-NOV-2001  (first entry)
XX
DE   Murine transport and binding associated protein encoding cDNA SEQ ID 676.
XX
KW   Murine; liver; gene library; amino acid synthesis; binding protein;
KW   cell metabolism; energy metabolism; fatty acid metabolism; synthesis;
KW   phospholipid metabolism; purine; pyrimidine; nucleoside; nucleotide;
KW   replication; transcription; translation; transport protein; ss.
XX
OS   Mus musculus.
XX
PN   DE20103510-U1.
XX
PD   07-JUN-2001.
XX
PF   28-FEB-2001; 2001DE-2003510.
XX
PR   02-DEC-1999;   99DE-1058160.
XX
PA   (LION-) LION BIOSCIENCE AG.
XX
DR   WPI; 2001-368570/39.
XX
PT   Gene library containing sequences with specific 3'-ends and no polyA
PT   tail, encoding proteins involved in a wide range of cellular processes
PT   -
XX
PS   Claim 15; Page 205-206; 251pp; German.
XX
CC   This invention describes a novel gene library (A) comprises a gene
CC   sequence (or its part) encoding a protein involved in amino acid
CC   synthesis, cellular/energy metabolism, metabolism of
CC   fatty acids/phospholipids, synthesis or breakdown of
CC   purines/pyrimidines/nucleosides/nucleotides, DNA
CC   replication/transcription/translation, or is a transport/binding protein.
CC   (A) are produced that correspond to the 3'-end of mRNA but without the
CC   polyA tail. They can be prepared more efficiently and with less effort
CC   than conventional libraries. AAK53436-AAK54275 represent fragments of the
CC   gene library described in the method of the invention.
XX
SQ   Sequence 513 BP; 143 A; 125 C; 107 G; 137 T; 1 other;

  Query Match             78.2%;  Score 17.2;  DB 22;  Length 513;
  Best Local Similarity   86.4%;  Pred. No. 1e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 AAGACTTCTGAGTAACAATCAA 22
           |||| ||||| |||||||| ||
Db      22 AAGATTTCTGGGTAACAATAAA 1


RESULT 6
ABL07736
ID   ABL07736 standard; cDNA; 7626 BP.
XX
AC   ABL07736;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 17690.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB63633.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
```

PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 17690; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7626 BP; 1991 A; 1673 C; 1791 G; 2171 T; 0 other;

Query Match 78.2%; Score 17.2; DB 23; Length 7626;
Best Local Similarity 86.4%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAGACTTCTGAGTAACAATCAA 22
|||| ||||||| |||||| ||
Db 5862 AAGAATTCTGAGAAACAATAAA 5883

RESULT 7
AAH68529
ID AAH68529 standard; DNA; 349980 BP.
XX
AC AAH68529;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 7064.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Disclosure; SEQ ID NO: 7064; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 349980 BP; 82466 A; 95954 C; 90516 G; 81044 T; 0 other;

Query Match 78.2%; Score 17.2; DB 22; Length 349980;
Best Local Similarity 86.4%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAGACTTCTGAGTAACAATCAA 22
|| ||||| || ||||||||||
Db 232860 AAAACTTCCGATTAACAATCAA 232881

RESULT 8
AAC57234
ID AAC57234 standard; DNA; 461 BP.
XX
AC AAC57234;
XX
DT 25-JAN-2001 (first entry)
XX
DE Pinus radiata transcription factor DNA sequence #631.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Pinus radiata.
XX
PN WO200053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06112.
XX
PR 11-MAR-1999; 99US-0266513.
PR 18-AUG-1999; 99US-0149485.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
DR WPI; 2000-579369/54.
XX
PT New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT -
XX
PS Claim 1; Page 624; 747pp; English.
XX
CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
XX

```
SQ   Sequence 461 BP; 163 A; 97 C; 94 G; 107 T; 0 other;

  Query Match             76.4%;  Score 16.8;  DB 21;  Length 461;
  Best Local Similarity   90.0%;  Pred. No. 1.6e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 GACTTCTGAGTAACAATCAA 22
           ||||||| ||||||| ||||
Db     408 GACTTCTCAGTAACAGTCAA 427


RESULT 9
AAK66626
ID   AAK66626 standard; DNA; 21477 BP.
XX
AC   AAK66626;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21438.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
```

PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 21438; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 21477 BP; 5311 A; 4999 C; 5256 G; 5911 T; 0 other;

Query Match 76.4%; Score 16.8; DB 22; Length 21477;
Best Local Similarity 90.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGACTTCTGAGTAACAATCA 21
|| |||||||||| ||||||
Db 7221 AGCCTTCTGAGTACCAATCA 7240

RESULT 10
AAK66625
ID AAK66625 standard; DNA; 21480 BP.
XX
AC AAK66625;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21437.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 21437; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 21480 BP; 5315 A; 5005 C; 5257 G; 5903 T; 0 other;

Query Match 76.4%; Score 16.8; DB 22; Length 21480;
Best Local Similarity 90.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGACTTCTGAGTAACAATCA 21
|| |||||||||| ||||||
Db 7222 AGCCTTCTGAGTACCAATCA 7241

RESULT 11
AAS01375/c
ID AAS01375 standard; cDNA; 3114 BP.
XX
AC AAS01375;
XX
DT 04-JUL-2001 (first entry)
XX
DE Human TANGO 405 cDNA sequence.
XX
KW Human; TANGO 210; clone jthLa152h06; TANGO 364; TANGO 366; dectin-2;
KW INTERCEPT 394; INTERCEPT 400; TANGO 405; cellular process regulator;
KW gene therapy; growth modulator; proliferation; cell differentiation;
KW lymphocyte; bone marrow cell migration; leukaemia; lymphoma;
KW autoimmune disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 154..783
FT /*tag= a
FT /product= "TANGO 405 protein"
FT /note= "The ORF is specifically claimed"
FT sig_peptide 154..297
FT /*tag= b
FT mat_peptide 298..780
FT /*tag= c
XX
PN WO200118016-A1.

XX
PD 15-MAR-2001.
XX
PF 30-JUN-2000; 2000WO-US18174.
XX
PR 10-SEP-1999; 99US-0393996.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Fraser CC, Sharp JD, Wrighton N, Myers PS, Goodearl ADJ;
XX
DR WPI; 2001-183280/18.
DR P-PSDB; AAU00479.
XX
PT Isolated nucleic acid molecules encoding proteins useful as modulating
PT agents in regulating a variety of cellular processes are used for
PT treating e.g. cancer and autoimmune disorders -
XX
PS Claim 2; Fig 6A-6C; 326pp; English.
XX
CC The present sequence encoding for human TANGO 405 protein is isolated
CC from cDNA clone jthLa152h06 from a human mixed lymphocyte reaction cDNA
CC library. It is 1 of 6 novel human proteins which include TANGO 210
CC (AAU00469), TANGO 364 (AAU00471), TANGO 366 (AAU00472), INTERCEPT 394
CC (AAU00473) and INTERCEPT 400 (AAU00476). Novel sequences for murine
CC TANGO 210 (AAU00470), INTERCEPT 400 (AAU00477), TANGO 405 (AAU00480) and
CC a rat INTERCEPT 400 (AAU00478) sequence are also described. The nucleic
CC acids encoding these novel proteins are useful as modulating agents in
CC regulating a variety of cellular processes and can be used to express
CC the proteins in a host cell in gene therapy applications. Human and
CC murine TANGO 405 proteins show sequence homology to murine dectin-2.
CC TANGO 405 modulates growth, proliferation, survival, differentiation,
CC activity, morphology and movement/migration of human lymphocytes and
CC bone marrow cells and tissues and can be used to prevent, diagnose or
CC treat leukaemia, lymphomas and autoimmune disorders.
XX
SQ Sequence 3114 BP; 1001 A; 527 C; 517 G; 1069 T; 0 other;

Query Match 74.5%; Score 16.4; DB 22; Length 3114;
Best Local Similarity 94.4%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 CTTCTGAGTAACAATCAA 22
||||||||||.|||||||
Db 1885 CTTCTGAGTATCAATCAA 1868

RESULT 12
ABA18026
ID ABA18026 standard; DNA; 29329 BP.
XX
AC ABA18026;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 10357.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 10357; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 29329 BP; 6864 A; 6756 C; 7225 G; 8484 T; 0 other;

Query Match 74.5%; Score 16.4; DB 22; Length 29329;
Best Local Similarity 94.4%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GACTTCTGAGTAACAATC 20
||||||||||.|||||||
Db 8110 GACTTCTGAGAAACAATC 8127

RESULT 13
ABA20511
ID ABA20511 standard; DNA; 29329 BP.
XX
AC ABA20511;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 12842.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 12842; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative

```
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 29329 BP; 6864 A; 6756 C; 7225 G; 8484 T; 0 other;

  Query Match             74.5%;  Score 16.4;  DB 22;  Length 29329;
  Best Local Similarity   94.4%;  Pred. No. 3.6e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 GACTTCTGAGTAACAATC 20
           |||||||||| |||||||
Db    8110 GACTTCTGAGAAACAATC 8127


RESULT 14
AAK70791
ID   AAK70791 standard; DNA; 29329 BP.
XX
AC   AAK70791;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25603.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
```

```
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 25603; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 29329 BP; 6864 A; 6756 C; 7225 G; 8484 T; 0 other;

  Query Match             74.5%;  Score 16.4;  DB 22;  Length 29329;
  Best Local Similarity   94.4%;  Pred. No. 3.6e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 GACTTCTGAGTAACAATC 20
           |||||||||| |||||||
Db    8110 GACTTCTGAGAAACAATC 8127


RESULT 15
AAK78512
ID   AAK78512 standard; DNA; 29329 BP.
XX
AC   AAK78512;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33324.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
```

PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 33324; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 29329 BP; 6864 A; 6756 C; 7225 G; 8484 T; 0 other;

Query Match 74.5%; Score 16.4; DB 22; Length 29329;
Best Local Similarity 94.4%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GACTTCTGAGTAACAATC 20
||||||||||.|||||||
Db 8110 GACTTCTGAGAAACAATC 8127

Search completed: March 17, 2003, 10:52:21
Job time : 217.253 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 10:23:12 ; Search time 777.688 Seconds
(without alignments)
458.154 Million cell updates/sec

Title: US-09-836-439-6
Perfect score: 22
Sequence: 1 aagacttctgagtaacaatcaa 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	18.4	83.6	347	9	AA970603	AA970603 op40f09.s
c	2	18.4	83.6	395	9	AA837289	AA837289 od26e07.s
c	3	17.8	80.9	364	9	AA229942	AA229942 nc51h06.r
	4	17.8	80.9	418	14	BQ622226	BQ622226 fch1c.pk0
c	5	17.8	80.9	509	12	BG882650	BG882650 sae94g07.
	6	17.8	80.9	545	9	AI738689	AI738689 wi22a08.x
c	7	17.8	80.9	580	9	AA546700	AA546700 vk55a10.r
c	8	17.8	80.9	857	17	BH138512	BH138512 ENTNI76TF
	9	17.4	79.1	375	12	BF609959	BF609959 NXSI_053_
c	10	17.4	79.1	474	10	AW699516	AW699516 gb06f02.y
	11	17.4	79.1	477	12	BG040453	BG040453 NXSI_108_
c	12	17.4	79.1	549	10	AW738913	AW738913 gb16b06.y
	13	17.4	79.1	608	17	AQ525332	AQ525332 HS_5226_B
	14	17.4	79.1	611	13	BJ173371	BJ173371 BJ173371
	15	17.4	79.1	692	13	BJ171523	BJ171523 BJ171523
	16	17.4	79.1	734	13	BJ167340	BJ167340 BJ167340
c	17	17.2	78.2	106	10	BE339863	BE339863 EST343923
c	18	17.2	78.2	185	17	AZ242233	AZ242233 RPCI-23-7
c	19	17.2	78.2	236	10	BB528073	BB528073 BB528073
c	20	17.2	78.2	258	10	BB301743	BB301743 BB301743
	21	17.2	78.2	325	17	AZ746982	AZ746982 RPCI-24-1
c	22	17.2	78.2	328	13	BI932924	BI932924 EST552813
c	23	17.2	78.2	390	12	BF451198	BF451198 uz66c12.x
c	24	17.2	78.2	391	12	BF323029	BF323029 maa38b04.
c	25	17.2	78.2	396	17	AZ104881	AZ104881 RPCI-23-4
c	26	17.2	78.2	409	12	BE851242	BE851242 uw93h01.y
	27	17.2	78.2	417	12	BE848177	BE848177 uw36b05.y
c	28	17.2	78.2	420	14	BM899883	BM899883 UI-M-DJ1-
	29	17.2	78.2	425	9	AI663816	AI663816 uj06c05.x
c	30	17.2	78.2	432	13	BI474888	BI474888 fp54a11.x
c	31	17.2	78.2	436	9	AU014847	AU014847 AU014847
c	32	17.2	78.2	450	14	BQ132786	BQ132786 fz68g09.x
	33	17.2	78.2	454	17	AQ769890	AQ769890 HS_3084_B
	34	17.2	78.2	460	10	BE579476	BE579476 kq28d05.y
c	35	17.2	78.2	473	10	AW220604	AW220604 EST296989
c	36	17.2	78.2	482	9	AI267625	AI267625 aq92b06.x
c	37	17.2	78.2	482	10	AW398852	AW398852 EST309352
c	38	17.2	78.2	487	10	BE462943	BE462943 EST324995
c	39	17.2	78.2	489	9	AU080451	AU080451 AU080451
c	40	17.2	78.2	489	9	AA270973	AA270973 va84c09.r
c	41	17.2	78.2	492	12	BF116785	BF116785 uz02c01.y
c	42	17.2	78.2	511	12	BF016374	BF016374 uy40b09.y
c	43	17.2	78.2	512	10	BE689588	BE689588 uw56d01.y
	44	17.2	78.2	515	12	BG359610	BG359610 sac26d02.
c	45	17.2	78.2	518	17	AQ416223	AQ416223 RPCI-11-1

ALIGNMENTS

RESULT 1
AA970603/c
LOCUS AA970603 347 bp mRNA linear EST 07-JUL-1998
DEFINITION op40f09.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1579337 3', mRNA sequence.
ACCESSION AA970603
VERSION AA970603.1 GI:3145110
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 347)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1053 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 344.
FEATURES Location/Qualifiers
source 1. .347
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1579337"

```
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT       91 a     75 c     63 g    118 t
ORIGIN

  Query Match             83.6%;  Score 18.4;  DB 9;  Length 347;
  Best Local Similarity   95.0%;  Pred. No. 4.1e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 GACTTCTGAGTAACAATCAA 22
           |||||||||||| |||||||
Db     335 GACTTCTGAGTACCAATCAA 316


RESULT 2
AA837289/c
LOCUS       AA837289                395 bp    mRNA    linear   EST 31-MAR-1998
DEFINITION  od26e07.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1369092,
            mRNA sequence.
ACCESSION   AA837289
VERSION     AA837289.1  GI:2912488
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1043   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 336.
FEATURES             Location/Qualifiers
     source          1. .395
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1369092"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      100 a     77 c     73 g    145 t
ORIGIN

  Query Match             83.6%;  Score 18.4;  DB 9;  Length 395;
  Best Local Similarity   95.0%;  Pred. No. 4.3e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 GACTTCTGAGTAACAATCAA 22
           |||||||||||| |||||||
Db     345 GACTTCTGAGTACCAATCAA 326


RESULT 3
AA229942/c
LOCUS       AA229942                364 bp    mRNA    linear   EST 21-AUG-1997
DEFINITION  nc51h06.r1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:1011707
            similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION   AA229942
VERSION     AA229942.1  GI:1852255
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
            , Michael Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 616   Std Error: 0.00
            Seq primer: -28m13 rev1 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .364
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1011707"
                     /clone_lib="NCI_CGAP_Pr3"
                     /sex="Male"
                     /dev_stage="45 years old"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; Site_1: Not1; Site_2: EcoRI; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNAse-treated, total cellular RNA obtained from 5,000-10
                     ,000 microdissected cells histologically-determined to be
                     fully malignant prostate cancer cells.  Double-stranded
                     cDNA was ligated to EcoRI adaptors, 5 cycles of PCR
                     applied to the cDNA with an adaptor-specific primer, and
                     the resulting PCR product subcloned into pAMP10 by the
                     UDG-cloning method (Life Technologies).  Average insert
                     size is 600 bp. NOTE: Not directionally cloned.  This
                     library was constructed by  David Krizman."
BASE COUNT      101 a     82 c     86 g     95 t
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 9;  Length 364;
  Best Local Similarity   90.5%;  Pred. No. 7.9e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy       2 AGACTTCTGAGTAACAATCAA 22
           || || ||||||||||||||||
Db      48 AGTCTGCTGAGTAACAATCAA 28
```

```
RESULT 4
BQ622226
LOCUS       BQ622226                418 bp    mRNA    linear   EST 01-JUL-2002
DEFINITION  fchlc.pk005.h2 Conidiobolus cornatus ARSEF 512 Conidiobolus
            coronatus cDNA, mRNA sequence.
ACCESSION   BQ622226
VERSION     BQ622226.1  GI:21649395
KEYWORDS    EST.
SOURCE      Conidiobolus coronatus.
  ORGANISM  Conidiobolus coronatus
            Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
            Ancylistaceae; Conidiobolus.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   Freimoser,F.M., Screen,S., Hu,G. and St. Leger,R.J.
  TITLE     EST analysis of genes expressed by the zygomycete pathogen
            Conidiobolus coronatus during optimized secretion of proteins
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Freimoser F. M.
            Department of Entomology
            Universtiy of Maryland
            4112 Plant Sciences Building, College Park, MD 20742, USA
            Tel: 301 405 16 13
            Fax: 301 314 92 90
            Email: ff34@umail.umd.edu.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Conidiobolus coronatus"
                     /strain="ARSEF 512"
                     /db_xref="taxon:34488"
                     /clone_lib="Conidiobolus cornatus ARSEF 512"
                     /note="Vector: UniZap; Conidiobolus coronatus was grown in
                     minimal medium supplemented with Manduca sexta cuticle and
                     peptone for 18 hours. A cDNA library was constructed in
                     the  unidirectional Lambda vector UniZap."
BASE COUNT      119 a     85 c     76 g    136 t      2 others
ORIGIN
```

```
  Query Match             80.9%;  Score 17.8;  DB 14;  Length 418;
  Best Local Similarity   90.5%;  Pred. No. 8.3e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 AGACTTCTGAGTAACAATCAA 22
           |||||| ||||||||| ||||
Db     175 AGACTTATGAGTAACACTCAA 195
```

```
RESULT 5
BG882650/c
LOCUS       BG882650                509 bp    mRNA    linear   EST 29-NOV-2001
DEFINITION  sae94g07.y1 Gm-c1065 Glycine max cDNA clone GENOME SYSTEMS CLONE
            ID: Gm-c1065-3758 5', mRNA sequence.
ACCESSION   BG882650
VERSION     BG882650.1  GI:14259742
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            High quality sequence stop: 415.
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1065-3758"
                     /clone_lib="Gm-c1065"
                     /tissue_type="germinating shoots"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     germinating shoots of the cultivar Williams. The seeds
                     were allowed to germinate for 24  hours prior to being
                     cold stressed for 2 days at 4C. Complementary DNA  was
                     synthesized from mRNA using a primer consisting of a
                     poly(dT)  sequence with a XhoI restriction site. EcoRI
                     adapters were ligated to  the blunt-ended cDNA fragments
                     followed by XhoI digestion. The cDNA  fragments were
                     directionally cloned into the EcoRI-XhoI restriction  site
                     of the pBluescript vector. The ligated cDNA fragments were
                     transformed into DH10B host cells (GibcoBRL). This library
                     was  constructed in the laboratory of Dr. Randy
                     Shoemaker."
BASE COUNT      151 a     91 c    121 g    146 t
ORIGIN
```

```
  Query Match             80.9%;  Score 17.8;  DB 12;  Length 509;
  Best Local Similarity   90.5%;  Pred. No. 8.9e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 AGACTTCTGAGTAACAATCAA 22
           || ||| ||||||||||||||||
Db      77 AGCCTTGTGAGTAACAATCAA 57
```

```
RESULT 6
AI738689
LOCUS       AI738689                545 bp    mRNA    linear   EST 20-DEC-1999
DEFINITION  wi22a08.x1 NCI_CGAP_Co16 Homo sapiens cDNA clone IMAGE:2390966 3'
            similar to TR:O54729 O54729 BRAIN. ;, mRNA sequence.
ACCESSION   AI738689
VERSION     AI738689.1  GI:5100670
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 545)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.
            , Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
```

```
           Insert Length: 850   Std Error: 0.00
           Seq primer: -40UP from Gibco
           High quality sequence stop: 446.
FEATURES             Location/Qualifiers
     source          1. .545
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2390966"
                     /clone_lib="NCI_CGAP_Co16"
                     /tissue_type="colon tumor, RER+"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Co10 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1057416-1061255, and 1144584-1145351).
                     Subtraction by Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT      165 a    131 c     82 g    167 t
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 9;  Length 545;
  Best Local Similarity   90.5%;  Pred. No. 9.1e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 AAGACTTCTGAGTAACAATCA 21
          |||| |||||||||| ||||||
Db    226 AAGAGTTCTGAGTATCAATCA 246


RESULT 7
AA546700/c
LOCUS       AA546700                 580 bp    mRNA    linear   EST 05-AUG-1997
DEFINITION  vk55a10.r1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone
            IMAGE:958554 5', mRNA sequence.
ACCESSION   AA546700
VERSION     AA546700.1  GI:2307991
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:547346
            High quality sequence stop: 405.
FEATURES             Location/Qualifiers
     source          1. .580
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:958554"
                     /clone_lib="Stratagene mouse Tcell 937311"
                     /tissue_type="Tcell"
                     /dev_stage="M30 CD4+ cells"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: blood; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. M30 CD4+ cells. Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      162 a    117 c    140 g    161 t
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 9;  Length 580;
  Best Local Similarity   90.5%;  Pred. No. 9.3e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 AGACTTCTGAGTAACAATCAA 22
          |||||||||| ||| ||||||
Db    506 AGACTTCTGACTAAGAATCAA 486


RESULT 8
BH138512/c
LOCUS       BH138512                 857 bp    DNA     linear   GSS 07-AUG-2001
DEFINITION  ENTNI76TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   BH138512
VERSION     BH138512.1  GI:15097573
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 857)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 8
            High quality sequence stop: 461.
FEATURES             Location/Qualifiers
     source          1. .857
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      276 a    151 c    101 g    329 t
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 17;  Length 857;
  Best Local Similarity   90.5%;  Pred. No. 1.1e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 AAGACTTCTGAGTAACAATCA 21
          |||||||||  ||||||||||
```

```
Db    509 AAGACTTCTTTGTAACAATCA 489


RESULT 9
BF609959
LOCUS       BF609959                375 bp    mRNA    linear   EST 14-DEC-2000
DEFINITION  NXSI_053_A10_F NXSI (Nsf Xylem Side wood Inclined) Pinus taeda cDNA
            clone NXSI_053_A10 5', mRNA sequence.
ACCESSION   BF609959
VERSION     BF609959.1  GI:11778131
KEYWORDS    EST.
SOURCE      loblolly pine.
  ORGANISM  Pinus taeda
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Sederoff,R.
  TITLE     Molecular Basis of Wood Formation in the Pine Megagenome
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Johnson, Arthur
            North Carolina State University
            Tel: 919 515 7800
            Fax: 919 515 7801
            Email: ajohnson@unity.ncsu.edu
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .375
                     /organism="Pinus taeda"
                     /strain="Coastal plain loblolly pine from North Carolina"
                     /db_xref="taxon:3352"
                     /clone="NXSI_053_A10"
                     /clone_lib="NXSI (Nsf Xylem Side wood Inclined)"
                     /tissue_type="Xylem"
                     /cell_type="Side"
                     /dev_stage="Juvenile"
                     /lab_host="XL1-Blue"
                     /note="Vector: BlueScript SK; Site_1: Eco RI; Site_2: XhoI
                     ; The library is from early (spring) wood, taken from
                     three six-year old trees (three different genotypes), in
                     the juvenile phase. These trees were induced to form side
                     wood by bending to a 45 degree angle and tying them to the
                     ground. Differentiating xylem was harvested from the sides
                     of the inclined stems, and a mixture of all three
                     genotypes was used for the library. oligo-dT primed cDNA
                     was directionally cloned into the EcoRI-XhoI BlueScript SK
                     vector arms. NOTE: The sequences contain a 'cDNA adapter'
                     between the EcoRI site and the start of the EST. The
                     adapter sequence is 'AATTCGGCACGAG'."
BASE COUNT       80 a     76 c     76 g    129 t     14 others
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 12;  Length 375;
  Best Local Similarity   94.7%;  Pred. No. 1.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 ACTTCTGAGTAACAATCAA 22
          ||||||||||||||| |||
Db    132 ACTTCTGAGTAACAACCAA 150


RESULT 10
AW699516/c
LOCUS       AW699516                474 bp    mRNA    linear   EST 18-APR-2000
DEFINITION  gb06f02.y1 Moss EST library PPN Physcomitrella patens cDNA clone
            PEP_SOURCE_ID:PPN080304 5' similar to TR:Q96538 Q96538 ACYL-COA
            SYNTHETASE ;, mRNA sequence.
ACCESSION   AW699516
VERSION     AW699516.1  GI:7583607
KEYWORDS    EST.
SOURCE      Physcomitrella patens.
  ORGANISM  Physcomitrella patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1  (bases 1 to 474)
  AUTHORS   Quatrano,R., Bashiardes,S., Cove,D., Cuming,A., Knight,C., Clifton
            ,S., Marra,M., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood
            ,K., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T.,
            Steptoe,M., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.
            , Waterston,R. and Wilson,R.
  TITLE     Leeds/Wash U Moss EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Ralph Quatrano
            Leeds/Wash U Moss EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Libraries were constructed by Dr. Stavros Bashiardes as part of the
            Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
            Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 405.
FEATURES             Location/Qualifiers
     source          1. .474
                     /organism="Physcomitrella patens"
                     /db_xref="taxon:3218"
                     /clone="PEP_SOURCE_ID:PPN080304"
                     /clone_lib="Moss EST library PPN"
                     /tissue_type="protonemata: 7 day old tissue auxin treated"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Construction of the cDNA library was carried out
                     using Stratagenes 'UniZAP - cDNA synthesis kit'.  cDNA was
                     constructed using an oligo dT primer/linker that contains
                     a XhoI site within it.  Following ds cDNA  synthesis,
                     EcoRI adapters were ligated to the blunt ends and sample
                     was digested with XhoI. The result is cDNA with an EcoRI
                     sticky end on one side and a XhoI sticky end on the other.
                     This cDNA was ligated directionally in UniZAP arms. The
                     vector is designed containing the pBluescript sequence as
                     well as lambda DNA and cDNA is cloned within this
                     pBluescript sequence. The vector was then packaged using
                     Gold gigapackaging extracts. Library was grown in XLIBlue
                     MRF' cells and amplified. The library was excised by mass
                     excision using Stratagens 'Mass excision kit' that uses
                     exassist as a helper phage that releases the pBluescript
                     sequence and circularises it as single stranded plasmids
                     that are then packaged (by helper phage) and secreted out
                     of the host cell as phagemids. SOLR cells were transformed
                     with phagemids and the library was plated out on LB-amp
                     plates to select for transformants. Approximately 1,000
                     ,000 colonies were grown and recovered. The double
                     stranded plasmid library was recovered by using Quiagen
                     Midi prep kit. 2 micro grams of each library were used to
                     transform DH10B cells by electroporation."
BASE COUNT      138 a     99 c    114 g    123 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 10;  Length 474;
  Best Local Similarity   94.7%;  Pred. No. 1.3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 ACTTCTGAGTAACAATCAA 22
          |||||||||||||| |||||
Db    432 ACTTCTGAGTAACCATCAA 414


RESULT 11
BG040453
LOCUS       BG040453                477 bp    mRNA    linear   EST 24-JAN-2001
DEFINITION  NXSI_108_D07_F NXSI (Nsf Xylem Side wood Inclined) Pinus taeda cDNA
            clone NXSI_108_D07 5', mRNA sequence.
ACCESSION   BG040453
```

```
VERSION     BG040453.1  GI:12483038
KEYWORDS    EST.
SOURCE      loblolly pine.
  ORGANISM  Pinus taeda
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Sederoff,R.
  TITLE     Molecular Basis of Wood Formation in the Pine Megagenome
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Johnson, Arthur
            North Carolina State University
            Tel: 919 515 7800
            Fax: 919 515 7801
            Email: ajohnson@unity.ncsu.edu
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .477
                     /organism="Pinus taeda"
                     /strain="Coastal plain loblolly pine from North Carolina"
                     /db_xref="taxon:3352"
                     /clone="NXSI_108_D07"
                     /clone_lib="NXSI (Nsf Xylem Side wood Inclined)"
                     /tissue_type="Xylem"
                     /cell_type="Side"
                     /dev_stage="Juvenile"
                     /lab_host="XL1-Blue"
                     /note="Vector: BlueScript SK; Site_1: Eco RI; Site_2: XhoI
                     ; The library is from early (spring) wood, taken from
                     three six-year old trees (three different genotypes), in
                     the juvenile phase. These trees were induced to form side
                     wood by bending to a 45 degree angle and tying them to the
                     ground. Differentiating xylem was harvested from the sides
                     of the inclined stems, and a mixture of all three
                     genotypes was used for the library. oligo-dT primed cDNA
                     was directionally cloned into the EcoRI-XhoI BlueScript SK
                     vector arms. NOTE: The sequences contain a 'cDNA adapter'
                     between the EcoRI site and the start of the EST. The
                     adapter sequence is 'AATTCGGCACGAG'."
BASE COUNT      106 a    106 c     93 g    154 t     18 others
ORIGIN

  Query Match            79.1%;  Score 17.4;  DB 12;  Length 477;
  Best Local Similarity  94.7%;  Pred. No. 1.3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 ACTTCTGAGTAACAATCAA 22
          |||||||||||||||| |||
Db    134 ACTTCTGAGTAACAACCAA 152


RESULT 12
AW738913/c
LOCUS       AW738913                549 bp    mRNA    linear   EST 25-APR-2000
DEFINITION  gb16b06.y1 Moss EST library PPN Physcomitrella patens cDNA clone
            PEP_SOURCE_ID:PPN100912 5' similar to TR:Q96538 Q96538 ACYL-COA
            SYNTHETASE ;, mRNA sequence.
ACCESSION   AW738913
VERSION     AW738913.1  GI:7647930
KEYWORDS    EST.
SOURCE      Physcomitrella patens.
  ORGANISM  Physcomitrella patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Quatrano,R., Bashiardes,S., Cove,D., Cuming,A., Knight,C., Clifton
            ,S., Marra,M., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood
            ,K., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T.,
            Steptoe,M., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.
            , Waterston,R. and Wilson,R.
  TITLE     Leeds/Wash U Moss EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Ralph Quatrano
            Leeds/Wash U Moss EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Libraries were constructed by Dr. Stavros Bashiardes as part of the
            Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
            Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 408.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Physcomitrella patens"
                     /db_xref="taxon:3218"
                     /clone="PEP_SOURCE_ID:PPN100912"
                     /clone_lib="Moss EST library PPN"
                     /tissue_type="protonemata: 7 day old tissue auxin treated"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Construction of the cDNA library was carried out
                     using Stratagenes 'UniZAP - cDNA synthesis kit'.  cDNA was
                     constructed using an oligo dT primer/linker that contains
                     a XhoI site within it.  Following ds cDNA  synthesis,
                     EcoRI adapters were ligated to the blunt ends and sample
                     was digested with XhoI. The result is cDNA with an EcoRI
                     sticky end on one side and a XhoI sticky end on the other.
                     This cDNA was ligated directionally in UniZAP arms. The
                     vector is designed containing the pBluescript sequence as
                     well as lambda DNA and cDNA is cloned within this
                     pBluescript sequence. The vector was then packaged using
                     Gold gigapackaging extracts. Library was grown in XLIBlue
                     MRF' cells and amplified. The library was excised by mass
                     excision using Stratagens 'Mass excision kit' that uses
                     exassist as a helper phage that releases the pBluescript
                     sequence and circularises it as single stranded plasmids
                     that are then packaged (by helper phage) and secreted out
                     of the host cell as phagemids. SOLR cells were transformed
                     with phagemids and the library was plated out on LB-amp
                     plates to select for transformants. Approximately 1,000
                     ,000 colonies were grown and recovered. The double
                     stranded plasmid library was recovered by using Quiagen
                     Midi prep kit. 2 micro grams of each library were used to
                     transform DH10B cells by electroporation."
BASE COUNT      151 a    116 c    125 g    157 t
ORIGIN

  Query Match            79.1%;  Score 17.4;  DB 10;  Length 549;
  Best Local Similarity  94.7%;  Pred. No. 1.4e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 ACTTCTGAGTAACAATCAA 22
          |||||||||||||| |||||
Db    418 ACTTCTGAGTAACCATCAA 400


RESULT 13
AQ525332
LOCUS       AQ525332                608 bp    DNA     linear   GSS 11-MAY-1999
DEFINITION  HS_5226_B1_C08_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=802 Col=15 Row=F, DNA sequence.
ACCESSION   AQ525332
VERSION     AQ525332.1  GI:4772652
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 608)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
```

```
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 802  row: F  column: 15
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 608.
FEATURES             Location/Qualifiers
     source          1. .608
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=802 Col=15 Row=F"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      221 a    123 c    107 g    147 t     10 others
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 17;  Length 608;
  Best Local Similarity   85.7%;  Pred. No. 1.4e+03;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 AAGACTTCTGAGTAACAATCA 21
           || |||| | ||||||||||
Db      22 AANACTTTTNAGTAACAATCA 42


RESULT 14
BJ173371
LOCUS       BJ173371                611 bp    mRNA    linear   EST 24-JAN-2002
DEFINITION  BJ173371 full length cDNA library, chloronemata and young
            gametophores Physcomitrella patens subsp. patens cDNA clone
            pph35g12 3', mRNA sequence.
ACCESSION   BJ173371
VERSION     BJ173371.1  GI:18341336
KEYWORDS    EST.
SOURCE      Physcomitrella patens subsp. patens.
  ORGANISM  Physcomitrella patens subsp. patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Fujita,T., Shin-i,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.
            , Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe
            ,M.
  TITLE     Comparison of the moss Physcomitrella patens genome with flowering
            plants genome
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp
            A backbone of the vector is basically from pBluescript(KS), that
            was in vivo excised from a modified lPS phage vector (Mo bi Tec,
            Germany). 5' end of the cDNA that was digested with XhoI was
            ligated to SalI site of the vector and the 3' end including polyA
            tail was ligated to BamHI site of the vector. cDNA instert could be
            amplified with conventional T7 and T3 primers. This full-length
            cDNA library was generated basically according to the method
            described in The Plant J 15, 707-720 (1998) Seki M. et al.
            Protonemata were blended by the POLYTRON, and then cultivated on
            the BCDATG medium for 13-14 days under the continuous light.
FEATURES             Location/Qualifiers
     source          1. .611
                     /organism="Physcomitrella patens subsp. patens"
                     /db_xref="taxon:145481"
                     /clone="pph35g12"
                     /clone_lib="full length cDNA library, chloronemata and
                     young gametophores"
                     /tissue_type="mixture of chloronemata and young
                     gametophores with 2 to 5 leaves"
BASE COUNT      178 a    135 c    128 g    170 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 13;  Length 611;
  Best Local Similarity   94.7%;  Pred. No. 1.4e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 ACTTCTGAGTAACAATCAA 22
           ||||||||||||| |||||
Db     181 ACTTCTGAGTAACCATCAA 199


RESULT 15
BJ171523
LOCUS       BJ171523                692 bp    mRNA    linear   EST 24-JAN-2002
DEFINITION  BJ171523 full length cDNA library, chloronemata and young
            gametophores Physcomitrella patens subsp. patens cDNA clone
            pph29d18 3', mRNA sequence.
ACCESSION   BJ171523
VERSION     BJ171523.1  GI:18339496
KEYWORDS    EST.
SOURCE      Physcomitrella patens subsp. patens.
  ORGANISM  Physcomitrella patens subsp. patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1  (bases 1 to 692)
  AUTHORS   Fujita,T., Shin-i,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.
            , Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe
            ,M.
  TITLE     Comparison of the moss Physcomitrella patens genome with flowering
            plants genome
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp
            A backbone of the vector is basically from pBluescript(KS), that
            was in vivo excised from a modified lPS phage vector (Mo bi Tec,
            Germany). 5' end of the cDNA that was digested with XhoI was
            ligated to SalI site of the vector and the 3' end including polyA
            tail was ligated to BamHI site of the vector. cDNA instert could be
            amplified with conventional T7 and T3 primers. This full-length
            cDNA library was generated basically according to the method
            described in The Plant J 15, 707-720 (1998) Seki M. et al.
            Protonemata were blended by the POLYTRON, and then cultivated on
            the BCDATG medium for 13-14 days under the continuous light.
FEATURES             Location/Qualifiers
     source          1. .692
                     /organism="Physcomitrella patens subsp. patens"
                     /db_xref="taxon:145481"
                     /clone="pph29d18"
```

/clone_lib="full length cDNA library, chloronemata and young gametophores"
/tissue_type="mixture of chloronemata and young gametophores with 2 to 5 leaves"

BASE COUNT 213 a 153 c 148 g 178 t
ORIGIN

Query Match 79.1%; Score 17.4; DB 13; Length 692;
Best Local Similarity 94.7%; Pred. No. 1.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 AAGACTTCTGAGTAACAAT 19
           ||||||||||| |||||||
Db     337 AAGACTTCTGACTAACAAT 355
```

Search completed: March 17, 2003, 13:09:28
Job time : 782.688 secs